

# **Modelling progression and heterogeneity in Alzheimer's disease**

*Alexandra Young*

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
of  
**University College London.**

Centre for Medical Image Computing  
Department of Computer Science  
University College London

March 30, 2017



I, Alexandra Young, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.





# Abstract

Alzheimer's disease is the most common form of dementia, accounting for approximately 60-80% of cases. Pathologically, the disease is characterised by the accumulation of amyloid plaques and neurofibrillary tangles in brain tissue, which give rise to downstream neurodegeneration and cognitive deficits. Biomarkers, such as volumetric measures of neurodegeneration derived from Magnetic Resonance Imaging, allow the progression of Alzheimer's disease to be monitored in vivo. Hypothetical models have been proposed that describe a distinct sequence of biomarker changes, but also heterogeneity in this sequence across different population subgroups. However, the quantitative evolution and heterogeneity of these biomarker changes has yet to be determined.

This thesis investigates the progression and heterogeneity of Alzheimer's disease by developing mathematical models of disease progression that characterise the evolution of biomarker measurements from cross-sectional data. Three key contributions are made. First, the application of data-driven models to sporadic and dominantly-inherited Alzheimer's disease to determine the sequence of biomarker changes in each form of Alzheimer's disease, and to ascertain the utility of patient staging systems derived from the models. Second, the development of a simulation framework that produces synthetic neurodegenerative disease datasets, allowing the evaluation of the performance of mathematical models of disease progression. Third, the formulation of a data-driven subtyping model that uniquely uncovers population subgroups with distinct biomarker trajectories, enabling the separation of disease subtype from disease stage. Application of this model to sporadic Alzheimer's disease provides a novel data-driven classification of Alzheimer's

disease into subtypes with distinct patterns of regional volume loss, as well as fine-grained subtyping and staging information.

The models proposed in this thesis have wide potential further application to advance disease understanding and to provide precise patient staging information for other diseases and developmental processes.

# Acknowledgements

There are numerous people I would like to thank for supporting me during my PhD. Foremost, my two supervisors, Danny and Jon, for their expert guidance and remarkable commitment to the project. The POND research group, for many excellent suggestions and interesting/entertaining discussions and coffee breaks. My collaborators in CMIC, the DRC, the rest of UCL and beyond, whose knowledge and assistance is fundamental to my research. MIG, for their helpful comments and exceptional cake selections. My parents, for their loving support and lighthearted spirit. My siblings, for never letting me forget that I am the nerdiest! My friends, especially my housemates and colleagues, who keep me entertained with their great sense of humour. Finally, my partner Jon, who always brightens my day.

I would also like to thank my examiners for their constructive feedback on the thesis, my funding body, the EPSRC, and all of those involved in the ADNI, DIAN and GENFI studies, without whom this research would not have been possible.



# Contents

<b>1</b>	<b>Introduction</b>	<b>23</b>
1.1	Alzheimer's disease . . . . .	23
1.1.1	Background . . . . .	23
1.1.2	History . . . . .	24
1.1.3	Pathology . . . . .	24
1.1.4	Genetics . . . . .	27
1.1.5	Mechanism . . . . .	28
1.1.6	Biomarkers . . . . .	31
1.1.7	Diagnostic criteria . . . . .	35
1.1.8	Treatments . . . . .	37
1.2	Research problem . . . . .	38
1.2.1	Problem statement . . . . .	38
1.2.2	Justification . . . . .	38
1.2.3	Motivation . . . . .	39
1.3	Thesis contributions . . . . .	39
1.3.1	Data-driven models of biomarker changes in sporadic and dominantly-inherited Alzheimer's disease . . . . .	39
1.3.2	A simulation system for biomarker evolution in neurode- generative disease . . . . .	40
1.3.3	A data-driven model of disease subtypes with distinct pat- terns of biomarker evolution in frontotemporal dementia and Alzheimer's disease . . . . .	41
1.4	Structure of this thesis . . . . .	42

<b>2</b>	<b>State of the art in Alzheimer's disease progression modelling</b>	<b>45</b>
2.1	Associated publications . . . . .	45
2.2	Introduction . . . . .	45
2.3	Neuropathological models . . . . .	46
2.3.1	Neurofibrillary tangle accumulation . . . . .	46
2.3.2	Amyloid plaque deposition . . . . .	47
2.3.3	Critical assessment . . . . .	48
2.4	Animal models . . . . .	49
2.4.1	Transgenic mouse models . . . . .	49
2.4.2	Knockout mouse models . . . . .	49
2.4.3	Critical assessment . . . . .	50
2.5	Scalar biomarker models . . . . .	50
2.5.1	Comparing diagnostic groups . . . . .	50
2.5.2	Using cut points to define biomarker abnormality . . . . .	50
2.5.3	Indexing by cognitive test scores . . . . .	51
2.5.4	Indexing by estimated years to onset . . . . .	51
2.5.5	Indexing by conversion between diagnoses . . . . .	52
2.5.6	Critical assessment . . . . .	53
2.6	High dimensional biomarker models . . . . .	54
2.6.1	Comparing diagnostic groups . . . . .	54
2.6.2	Prediction of conversion between diagnoses . . . . .	55
2.6.3	Network models . . . . .	55
2.6.4	Critical assessment . . . . .	56
2.7	Data-driven biomarker models . . . . .	56
2.7.1	The event-based model . . . . .	57
2.7.2	Differential equation models . . . . .	57
2.7.3	Self-modelling regression . . . . .	58
2.7.4	Critical assessment . . . . .	60
2.8	Summary . . . . .	61

<b>3</b>	<b>A data-driven model of biomarker changes in sporadic Alzheimer's disease</b>	<b>63</b>
3.1	Associated publications . . . . .	63
3.2	Introduction . . . . .	63
3.3	Materials and methods . . . . .	65
3.3.1	Data description . . . . .	65
3.3.2	Event sequences . . . . .	66
3.3.3	Patient staging . . . . .	70
3.3.4	Staging using cross-sectional data alone . . . . .	71
3.4	Results . . . . .	72
3.4.1	Subjects . . . . .	72
3.4.2	Event sequences . . . . .	72
3.4.3	Patient staging . . . . .	75
3.4.4	Staging using cross-sectional data alone . . . . .	79
3.5	Discussion . . . . .	84
3.5.1	Event sequence . . . . .	85
3.5.2	Patient staging . . . . .	89
3.5.3	Model assumptions . . . . .	90
3.6	Conclusion . . . . .	92
<b>4</b>	<b>A data-driven model of biomarker changes in dominantly-inherited Alzheimer's disease</b>	<b>93</b>
4.1	Associated publications . . . . .	93
4.2	Introduction . . . . .	94
4.3	Materials and methods . . . . .	94
4.3.1	Data description . . . . .	94
4.3.2	Event-based models . . . . .	97
4.3.3	Implementation . . . . .	98
4.4	Results . . . . .	98
4.4.1	Event sequences . . . . .	98
4.4.2	Patient staging . . . . .	100

4.5	Discussion . . . . .	103
4.5.1	Event sequences . . . . .	103
4.5.2	Patient staging . . . . .	104
4.6	Conclusion . . . . .	104
<b>5</b>	<b>A simulation system for biomarker evolution in neurodegenerative dis-</b>	
	<b>ease</b>	<b>105</b>
5.1	Associated publications . . . . .	105
5.2	Introduction . . . . .	105
5.3	Materials and methods . . . . .	107
5.3.1	Generative model of data . . . . .	107
5.3.2	Simulating sporadic Alzheimer's disease . . . . .	107
5.3.3	The event-based model . . . . .	115
5.3.4	Differential equation model . . . . .	116
5.3.5	Evaluation metrics . . . . .	116
5.3.6	Implementation . . . . .	118
5.4	Results . . . . .	118
5.4.1	Event-based model stability analysis . . . . .	118
5.4.2	Differential equation model stability analysis . . . . .	123
5.5	Discussion . . . . .	125
5.5.1	Simulation framework . . . . .	125
5.5.2	Stability analysis . . . . .	126
5.5.3	Implications for the application and development of data- driven models . . . . .	127
5.6	Conclusion . . . . .	128
<b>6</b>	<b>Multiple orderings of events in disease progression</b>	<b>129</b>
6.1	Associated publications . . . . .	129
6.2	Introduction . . . . .	129
6.3	Models . . . . .	131
6.3.1	The event-based model . . . . .	131



6.3.2	The generalised Mallows event-based model . . . . .	132
6.3.3	Dirichlet process mixtures of generalised Mallows event-based models . . . . .	132
6.4	Inference . . . . .	133
6.4.1	The event-based model . . . . .	133
6.4.2	The generalised Mallows event-based model . . . . .	133
6.4.3	Dirichlet process mixtures of generalised Mallows event-based models . . . . .	134
6.5	Implementation . . . . .	134
6.5.1	ADNI dataset . . . . .	134
6.5.2	Model fitting . . . . .	135
6.6	Results and discussion . . . . .	136
6.6.1	The event-based model . . . . .	136
6.6.2	The generalised Mallows event-based model . . . . .	137
6.6.3	Dirichlet process mixtures of generalised Mallows event-based models . . . . .	138
6.7	Conclusion . . . . .	140
<b>7</b>	<b>A data-driven model of disease subtypes with distinct patterns of biomarker evolution in frontotemporal dementia and Alzheimer's disease</b>	<b>143</b>
7.1	Introduction . . . . .	143
7.2	Methods . . . . .	147
7.2.1	Data description . . . . .	147
7.2.2	Mathematical modelling overview . . . . .	149
7.2.3	Mathematical model . . . . .	151
7.2.4	Model fitting . . . . .	152
7.2.5	Implementation . . . . .	153
7.3	Results . . . . .	155
7.3.1	Genetic frontotemporal dementia . . . . .	155
7.3.2	Alzheimer's disease . . . . .	162

7.4	Discussion . . . . .	165
7.4.1	Genetic frontotemporal dementia . . . . .	166
7.4.2	Alzheimer's disease . . . . .	169
7.5	Conclusion . . . . .	170
<b>8</b>	<b>Further work</b>	<b>173</b>
8.1	Overview . . . . .	173
8.2	Further applications . . . . .	173
8.2.1	Application in neurodegenerative diseases . . . . .	173
8.2.2	Application to other diseases or processes . . . . .	175
8.3	Methodological developments of the models presented in this thesis	175
8.3.1	Event distributions . . . . .	175
8.3.2	Partial sequences . . . . .	176
8.3.3	Time . . . . .	176
8.3.4	Missing data . . . . .	177
8.3.5	High-dimensional biomarkers . . . . .	177
8.3.6	Within-subject models . . . . .	178
8.4	Broader technical advances . . . . .	178
8.4.1	Mixed pathology . . . . .	178
8.4.2	Integration of data-driven and mechanistic models . . . . .	178
8.4.3	Dimensionality reduction . . . . .	179
8.4.4	Incorporating additional data types . . . . .	179
8.4.5	Validation . . . . .	180
8.4.6	Patient staging systems . . . . .	180
8.4.7	Biomarker utility . . . . .	181
8.5	Summary . . . . .	181
<b>9</b>	<b>Summary and conclusion</b>	<b>183</b>
9.1	Summary of Chapter 1 . . . . .	183
9.2	Summary of Chapter 2 . . . . .	184
9.3	Summary of Chapter 3 . . . . .	184

9.4	Summary of Chapter 4 . . . . .	186
9.5	Summary of Chapter 5 . . . . .	186
9.6	Summary of Chapter 6 . . . . .	187
9.7	Summary of Chapter 7 . . . . .	188
9.8	Summary of Chapter 8 . . . . .	189
9.9	Conclusion . . . . .	190
<b>Appendices</b>		<b>191</b>
<b>A</b>	<b>Supplementary figures and tables for Chapter 7</b>	<b>191</b>
<b>Bibliography</b>		<b>201</b>



# List of Figures

1.1	Diagram of a neuron and neuroglial cells. . . . .	25
1.2	Phases of $\beta$ -amyloidosis proposed by Thal et al. in 2002. . . . .	26
1.3	Neurofibrillary tangle staging system proposed by Braak and Braak in 1991. . . . .	27
1.4	The amyloid cascade hypothesis. . . . .	29
1.5	The possible role of different Alzheimer's disease genes in $A\beta$ - related pathogenic processes. . . . .	30
1.6	Hypothetical model of Alzheimer's disease as a complex interaction between cellular, molecular and genetic imbalances. . . . .	32
1.7	Hypothetical model of Alzheimer's disease biomarker progression. .	34
2.1	Biomarker changes in dominantly-inherited Alzheimer's disease es- timated by Bateman et al. in 2012 by regressing against expected years to onset. . . . .	53
2.2	Sequence of events in dominantly-inherited Alzheimer's disease es- timated by the event-based model proposed by Fonteijn et al. in 2012.	58
2.3	Differential equation model of amyloid deposition proposed by Villemagne et al. in 2013. . . . .	59
2.4	Alzheimer's disease biomarker trajectories estimated using the self- modelling regression technique proposed by Donohue et al. in 2014.	60
3.1	Positional variance diagrams showing the distribution of event se- quences in various population subgroups. . . . .	75

3.2	Proportion of patients in each diagnostic category at each event-based model stage. . . . .	76
3.3	Longitudinal consistency of patient staging in the whole population. . . . .	77
3.4	Estimated probability of remaining with a MCI or a CN diagnosis for different baseline event-based model stages. . . . .	80
3.5	As Figure 3.1, but without using atrophy rates. . . . .	82
3.6	As Figure 3.2, but without using atrophy rates. . . . .	83
3.7	As Figure 3.3, but without using atrophy rates. . . . .	83
3.8	As Figure 3.4, but without using atrophy rates. . . . .	86
4.1	Sequence of biomarker changes estimated by the event-based model for dominantly-inherited Alzheimer's disease. . . . .	99
4.2	Sequence of biomarker changes estimated by the event-based model for APOE4-positive and APOE4-negative dominantly-inherited Alzheimer's disease mutation carriers. . . . .	100
4.3	Sequence of biomarker changes estimated by the event-based model for PSEN1, PSEN2, and APP mutation carriers. . . . .	101
4.4	Patient staging using the event-based model for dominantly-inherited Alzheimer's disease. . . . .	102
5.1	Generative model of sporadic disease datasets used by the simulation system. . . . .	108
5.2	Example simulated FDG-PET, subject diagnosis, and time point data generated using default parameter values for the event-based model. . . . .	114
5.3	Results of applying the event-based model to synthetic data with missing values and without missing values, generated using the default parameters. . . . .	119
5.4	Integrated differential equation model trajectories for FDG-PET uptake generated using synthetic data with the default settings for the event-based model. . . . .	124

6.1	Central ordering estimated by the event-based model. . . . .	137
6.2	Key for Figures 6.3 and 6.5. . . . .	138
6.3	Comparison of the central ordering estimated by the event-based model with the generalised Mallows model. . . . .	139
6.4	Estimate of the uncertainty in a subject's disease stage obtained by using Gibbs sampling to fit the generalised Mallows event-based model. . . . .	140
6.5	As Figure 6.3, but for the clusters identified by the Dirichlet process mixture of generalised Mallows event-based models. . . . .	141
7.1	Dynamic clustering of all mutation carriers in GENFI. . . . .	156
7.2	Fitting a single dynamic cluster to each mutation type in GENFI separately. . . . .	157
7.3	Dynamic clustering of C9orf72 mutation carriers. . . . .	158
7.4	Static clustering of all mutation carriers in GENFI. . . . .	161
7.5	Dynamic clustering of 3T ADNI dataset. . . . .	164
A.1	Dynamic clustering of all mutation carriers in GENFI: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. . . . .	192
A.2	Model likelihood for each of the dynamic cluster models fitted to all mutation carriers in GENFI. . . . .	193
A.3	Fitting a single dynamic cluster to each mutation type separately: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. . . . .	193
A.4	As Figure A.2, but for GRN mutation carriers. . . . .	194
A.5	Dynamic clustering of MAPT mutation carriers. . . . .	194
A.6	Dynamic clustering of MAPT mutation carriers: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. . . . .	195
A.7	As Figure A.2, but for MAPT mutation carriers. . . . .	196

A.8	Dynamic clustering of C9orf72 mutation carriers: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. . . . .	196
A.9	As Figure A.2, but for C9orf72 mutation carriers. . . . .	197
A.10	Dynamic clustering of 3T ADNI dataset: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. . . . .	197
A.11	As Figure A.2, but for 3T ADNI dataset. . . . .	198
A.12	Dynamic clustering of 1.5T ADNI dataset. . . . .	198
A.13	Dynamic clustering of 1.5T ADNI dataset: MCMC samples of uncertainty on top two rows, 10-fold cross-validation on bottom two rows. . . . .	199
A.14	As Figure A.2, but for 1.5T ADNI dataset. . . . .	200



# List of Tables

3.1	Baseline demographics for the whole population and population subgroups. . . . .	72
3.2	Classification results for discriminating MCI-stable versus MCI-converters and CN-stable versus CN-converters using patient stage at baseline. . . . .	78
3.3	Hazard ratios with 95% confidence intervals for conversion from MCI to AD, and CN to MCI. . . . .	79
3.4	As Table 3.1, but without using atrophy rates. . . . .	81
3.5	As Table 3.2, but without using atrophy rates. . . . .	84
3.6	As Table 3.3, but without using atrophy rates. . . . .	85
3.7	Cut point values derived using the event distributions estimated by the event-based model. . . . .	90
4.1	Demographics of Dominantly Inherited Alzheimer Network participants included in Data Freeze 6. . . . .	96
5.1	Default parameter values for the event-based model. . . . .	113
5.2	Mean Kendall's tau distance between ground truth event sequence and event sequence returned by the event-based model for synthetic data with various simulation settings. . . . .	121
5.3	Mean estimated transition time for differential equation model fitted to synthetic data with varying simulation settings. . . . .	125

7.1	Proportion of affected mutation carriers in GENFI assigned to each dynamic cluster using (A) in-sample models, and (B) out-of-sample models obtained from 10-fold cross-validation. . . . .	160
7.2	Proportion of affected mutation carriers in GENFI assigned to each static cluster using (A) in-sample models, and (B) out-of-sample models obtained from 10-fold cross-validation. . . . .	161
7.3	Hazards ratios for risk of conversion from MCI to AD in ADNI 3T dataset. . . . .	165
A.1	Parameters for each of the static clusters shown in Figure 7.4 . . . .	191

## **Chapter 1**

# **Introduction**

## **1.1 Alzheimer's disease**

### **1.1.1 Background**

Alzheimer's disease (AD) is a progressive neurodegenerative disease that is characterised by the build up of amyloid plaques and neurofibrillary tangles in brain tissue. It is the most common cause of dementia, accounting for approximately 60-80% of cases [1–3], with the term dementia referring to the group of symptoms caused by various brain disorders, rather than the underlying condition. The symptoms of dementia include memory loss, language difficulties, loss of problem solving abilities, behavioural problems, and difficulty performing activities of daily living. In the advanced disease stages there is a loss of bodily functions, which ultimately becomes fatal. Each different type of dementia is associated with a particular set, severity, and sequence of symptoms. AD typically presents with short-term memory loss, with a broader range of symptoms developing as the disease progresses, such as impaired communication, disorientation, confusion, and behavioural issues. There is however considerable overlap between the symptoms of different dementias, which, together with the possible presence of mixed dementia, complicates differential diagnosis. In addition to the devastating personal impact of dementia, the economic impact is huge, costing an estimated US \$818 billion worldwide [4]. The number of people living with dementia globally was estimated to be 46.8 million in 2015, with this number being predicted to almost double every 20 years [4]. Currently there

are no disease modifying treatments available; treatments only mask the symptoms of dementia temporarily.

### 1.1.2 History

In 1906 Alois Alzheimer presented the case of Auguste D. [5], a woman admitted to a mental institution at the age of 51 with progressive cognitive impairment that included memory problems and hallucinations [6]. Autopsy examination of her brain revealed the presence of plaques and neurofibrillary tangles [6, 7]. This particular type of dementia became known as Alzheimer's disease following the introduction of the eponym by Emil Kraepelin in 1910 [6, 8].

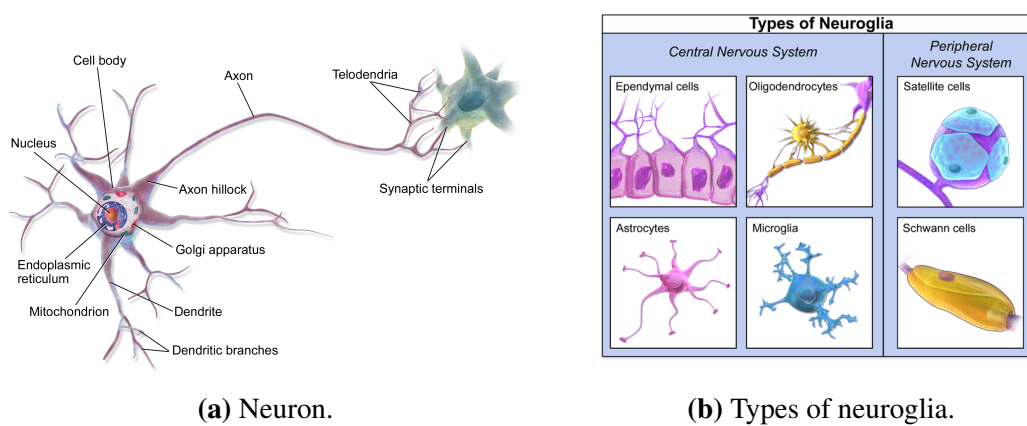
AD was initially thought to be a rare form of presenile dementia [6]; it wasn't until 1968 that it was shown that the neuropathology of many common senile dementia cases was indistinguishable from that of AD [9], leading to widespread recognition that AD is in fact a very common brain disorder [10]. From 1975 there has been a surge in AD research, with the number of publications increasing exponentially each year [11]. The National Institute of Aging, founded in 1974, and the Alzheimer's Association, founded in 1979, greatly increased the scientific and public awareness of AD [10], as did Robert Katzman's editorial in 1976 [12], which brought attention to the disease as an impending public health challenge [10].

### 1.1.3 Pathology

$A\beta$  plaques (amyloid plaques) and neurofibrillary tangles are the hallmark pathologies that define AD. They are thought to cause damage to neurons (Figure 1.1a) and neuroglial cells (Figure 1.1b) in the brain, although their precise mechanism of action remains unclear [13]. Neurons transmit information in the brain; neuroglia provide support for neurons.

#### 1.1.3.1 $A\beta$ plaques

$A\beta$  plaques are extracellular deposits of  $A\beta$ , which can broadly be classified as dense-core (neuritic, meaning that they consist of deteriorating neuronal material) or diffuse [14]. Dense-core plaques tend to be surrounded by dystrophic neurites (axons or dendrites - parts of neurons - shown in Figure 1.1a), reactive astrocytes



**Figure 1.1:** Diagram of a neuron<sup>1</sup> and neuroglial cells<sup>2</sup>.

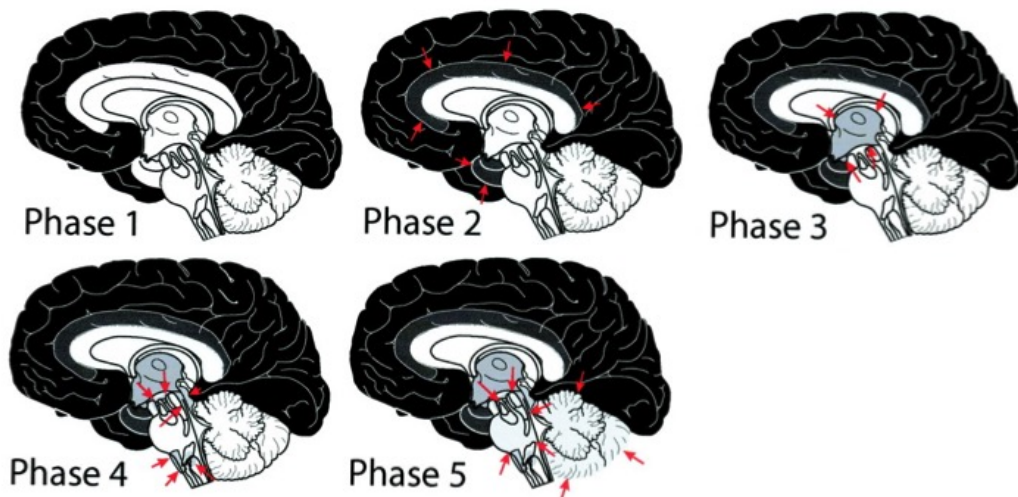
and activated microglial cells (types of neuroglia - see Figure 1.1b), and are associated with synaptic loss [15–18]. Diffuse plaques are usually non-neuritic and are not associated with glial activation or synaptic loss [14]. The presence of dense-core plaques is necessary for the pathological diagnosis of AD [19], whereas diffuse plaques are relatively common in cognitively normal elderly people [20–22]. The spatiotemporal pattern of amyloid deposition varies, but general patterns have been proposed that can be used to stage patients, such as that of Thal et al. in 2002 (Figure 1.2) [23].

### 1.1.3.2 Neurofibrillary tangles

Neurofibrillary tangles are intraneuronal aggregates of hyperphosphorylated and misfolded tau [14]. They become extraneuronal following the death of tangle-containing neurons [24,25]. Neuropil threads accompany neurofibrillary tangles, which are segments of axons and dendrites that contain aggregated and hyperphosphorylated tau [24]. The spatiotemporal pattern of neurofibrillary tangle progression is generally quite homogeneous, as described by Braak and Braak in 1991 (Figure 1.3) [26]. However, there are atypical cases. In 2011 Murray et al. [27] proposed that there are three subtypes of AD based on neurofibrillary tangle distribution: typical AD, hippocampal sparing AD, and limbic-predominant AD, with typical AD

<sup>1</sup>Image from User:BruceBlaus at Wikimedia Commons. CC BY 3.0.

<sup>2</sup>Image from Blausen.com staff (2014). “Medical gallery of Blausen Medical 2014”. WikiJournal of Medicine 1 (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. CC BY 3.0.



**Figure 1.2:** Phases of  $\beta$ -amyloidosis proposed by Thal et al. in 2002 [23]. In phase 1 there are  $A\beta$  deposits in the frontal, parietal, temporal, or occipital neocortex (shown in black). For each phase after phase 1 the red arrows indicate the presence of new  $A\beta$  deposits. Reprinted from [23], Copyright 2002, with permission from Wolters Kluwer Health, Inc.

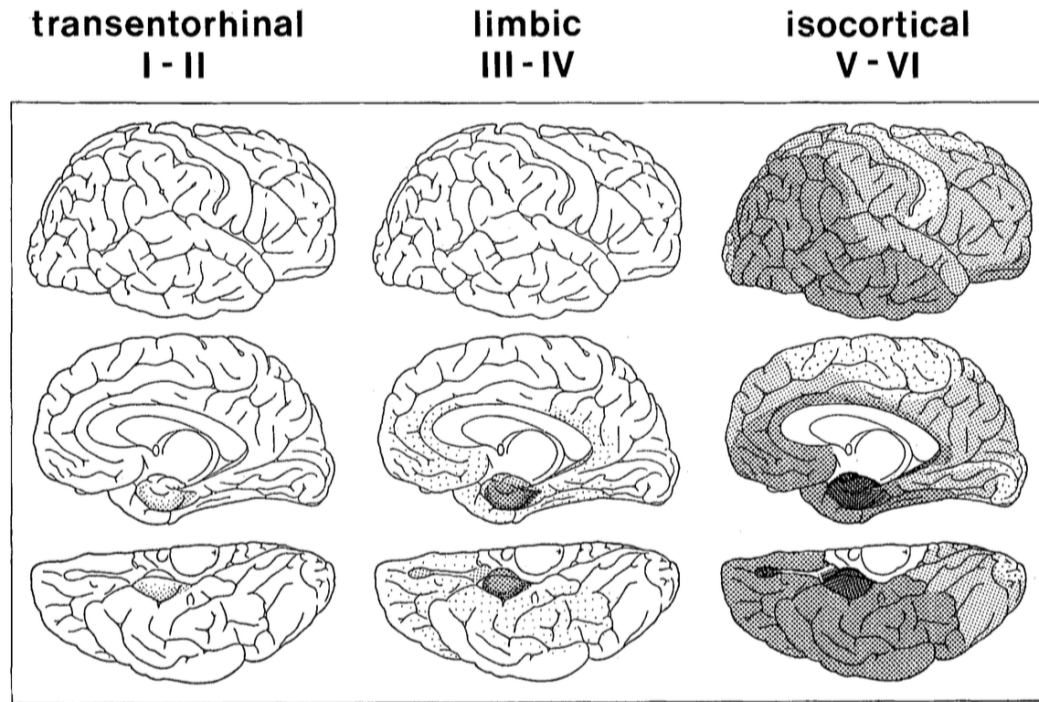
accounting for around 75% of cases.

### 1.1.3.3 Co-occurring pathologies

Several other pathologies commonly co-occur in AD: cerebral amyloid angiopathy, where amyloid deposits accumulate on the walls of cortical blood vessels; granulo-vacuolar degeneration, consisting of intraneuronal clusters of small vacuoles found in the hippocampal pyramidal neurons; Hirano bodies, which are perineuronal lesions found in the hippocampal CA1 region; glial responses in the form of reactive astrocytes and activated microglial cells; neuronal and synapse loss [14].

### 1.1.3.4 Associated conditions

Other conditions commonly coexist with AD, such as vascular dementia and Parkinson's disease. Consequently, non AD neuropathological lesions, e.g. ischemic infarctions and Lewy bodies, are frequently found alongside AD lesions. In the presence of mixed pathology it is difficult to disentangle the relative contribution of each condition to cognitive impairment [31].



**Figure 1.3:** Neurofibrillary tangle staging system proposed by Braak and Braak in 1991 [26]. The density of the shading indicates the severity of the neurofibrillary changes. Reprinted from [26], Copyright 1991, with permission from Springer.

### 1.1.4 Genetics

AD consists of two forms: an early onset dominantly-inherited form, which typically begins before the age of 60, and a late onset sporadic form, for which disease risk increases with age [32].

#### 1.1.4.1 Dominantly-inherited Alzheimer's disease

Mutations in three genes: the amyloid precursor protein (APP) gene [33], the presenillin 1 (PSEN1) gene [34], and the presenillin 2 (PSEN2) gene [35, 36], are known to cause early onset AD. All but one of the  $\sim 200$  pathogenic mutations in these three genes are autosomal-dominant [32], meaning that if you inherit one copy of the gene you will get the disease. The consequence of this is that a parent with early onset AD has a 50% chance of passing the condition on to their children. Most pathogenic AD mutations lead to a common molecular phenotype: an increase in the ratio of  $A\beta_{42}$  to  $A\beta_{40}$  [37].  $A\beta_{40}$  and  $A\beta_{42}$  are the two most common isoforms of  $A\beta$ , of which  $A\beta_{42}$  is considered to be the most associated with disease states.

### 1.1.4.2 Sporadic Alzheimer's disease

Late onset AD is more complex, with disease risk likely resulting from the interaction of a variety of genetic, environmental and lifestyle factors [32]. The most well established risk gene for AD is the apolipoprotein E (APOE) gene [38]. It is thought that APOE modifies disease risk by influencing  $A\beta$  metabolism. APOE has three possible isoforms:  $\epsilon 2$ ,  $\epsilon 3$  and  $\epsilon 4$ , of which  $\epsilon 3$  is the most common. The  $\epsilon 4$ -allele of APOE increases the risk of AD, whereas the  $\epsilon 2$ -allele is protective [39]. Carrying either one or two APOE  $\epsilon 4$ -alleles increases disease risk 3-fold, or 8- to 10-fold, respectively [32]. Many other genes have been implicated in AD, but none of them have strong effects [32].

## 1.1.5 Mechanism

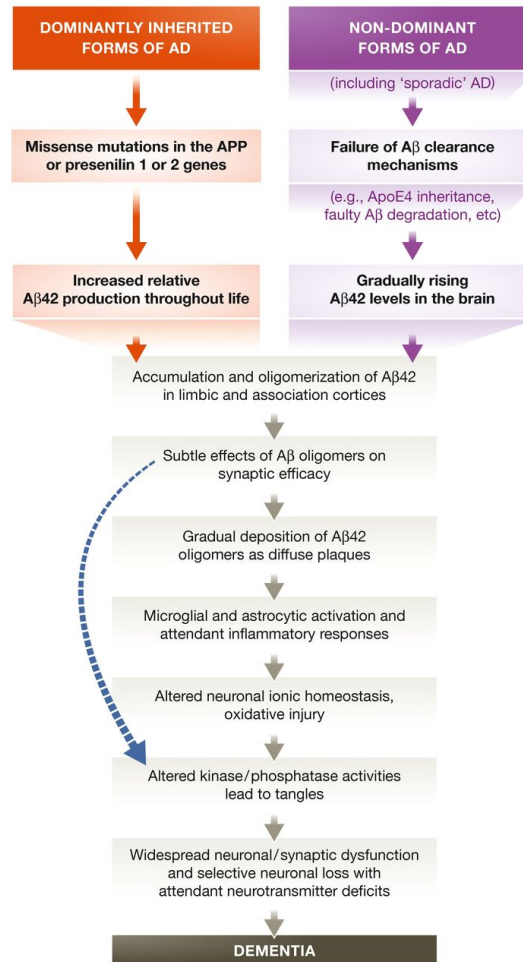
The biological mechanisms that give rise to the pathological hallmarks of AD, amyloid plaques and neurofibrillary tangles, are not well understood. It remains to be determined how the different pathologies propagate and interact, and how they give rise to downstream neurodegeneration and cognitive deficits.

### 1.1.5.1 The amyloid hypothesis

The amyloid hypothesis (Figure 1.4), which postulates that  $A\beta$  is the causative agent in AD, was first formalised in 1991 and 1992 [40–43] following the discovery of fully penetrant pathogenic mutations in the APP gene [33]. These mutations alter APP metabolism such that relatively more  $A\beta_{42}$  is produced [37,44], demonstrating that AD can result from an increase in the production of  $A\beta_{42}$ . However, 25 years later, the amyloid hypothesis is still a contentious subject [13].

There is strong evidence that  $A\beta$  is the initiating event in dominantly-inherited AD: mutations in the three major dominantly-inherited AD genes, APP, PSEN1, and PSEN2, all cause a relative increase in  $A\beta_{42}$  production [37]. Moreover, these three genes are all mechanistically involved in the production of  $A\beta$ : APP encodes the amyloid precursor protein from which  $A\beta$  is generated; PSEN1 and PSEN2 encode presenillin 1 and 2, which are subunits of the  $\beta$ - and  $\gamma$ -secretase complex that generates  $A\beta$  by cleaving APP. The potential role of these genes and other AD

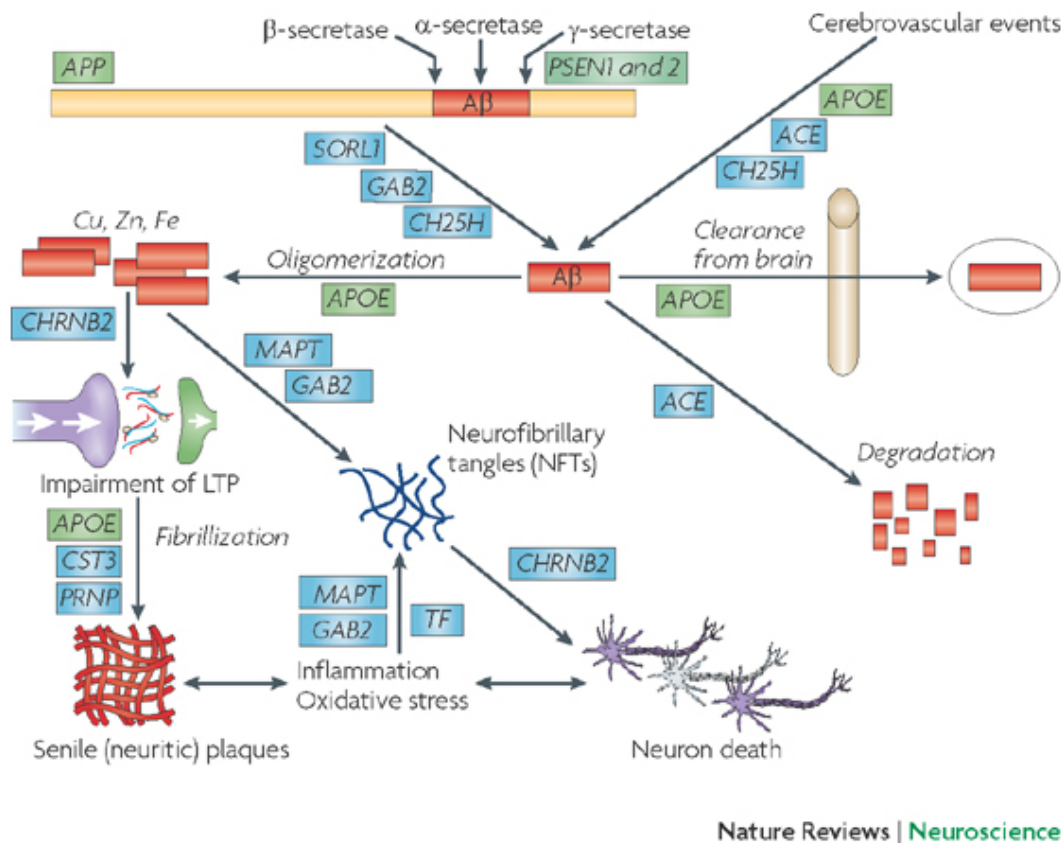




**Figure 1.4:** The amyloid cascade hypothesis. Reproduced from [13]. CC BY 4.0.

genes in A $\beta$ -related pathogenesis is summarised in Figure 1.5. There is however some debate as to whether A $\beta$ 42 itself contributes to dominantly-inherited AD, or whether the pathogenic mechanism is another process related to APP and its processing by presenellin [45]. The applicability of the amyloid cascade hypothesis to sporadic AD is disputed because of the genetic heterogeneity of the disease, but the major risk allele for sporadic AD, APOE  $\epsilon$ 4, has been shown to decrease A $\beta$  clearance, leading to a build up of A $\beta$  [46].

Another major argument in favour of the amyloid hypothesis is that amyloid plaques are more specific to AD than neurofibrillary tangles, which are involved in a range of tauopathies. However, the temporal and anatomical discord of amyloid plaques and tau pathology and neurodegeneration is difficult to explain [48]. Tem-



**Figure 1.5:** The possible role of different Alzheimer's disease genes in  $A\beta$ -related pathogenic processes. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience [47], Copyright 2008.

porally, biomarker studies show that amyloid plaque deposition begins long before tau deposition and neurodegeneration [49]. Anatomically, tau pathology correlates closely with neuronal loss, but amyloid plaque deposition does not correlate well with either [14, 29, 50].

A further concern is that the pathogenic mechanism of  $A\beta$  has yet to be determined, as has the function of  $A\beta$  and APP in the brain [13]. Recent versions of the amyloid hypothesis propose that the pathogenic process might be the formation of soluble  $A\beta_{42}$  oligomers [13]. Oligomeric  $A\beta$  may have subtle effects on the efficacy of synapses by decreasing synapse function and number, and may initiate tauopathy.

Critics of the amyloid hypothesis postulate that AD is a complex multifactorial disease that is not initiated by  $A\beta$  in isolation, with the presence of amyloid being

a necessary but not a sufficient cause [45].

#### 1.1.5.2 Tau and neurodegeneration

Evidence strongly suggests that dysfunction of tau causes neurodegeneration, however the exact mechanism by which this happens remains unclear. Dominantly inherited mutations in the MAPT gene [51–53], which encodes tau, cause frontotemporal dementia (FTD), demonstrating that dysfunction of tau is sufficient to cause neurodegeneration. Additionally neurofibrillary tangles, which are intraneuronal aggregates of hyperphosphorylated and misfolded tau, show better correlation with neurodegeneration than amyloid plaques [14, 29, 50].

#### 1.1.5.3 Propagation of pathology

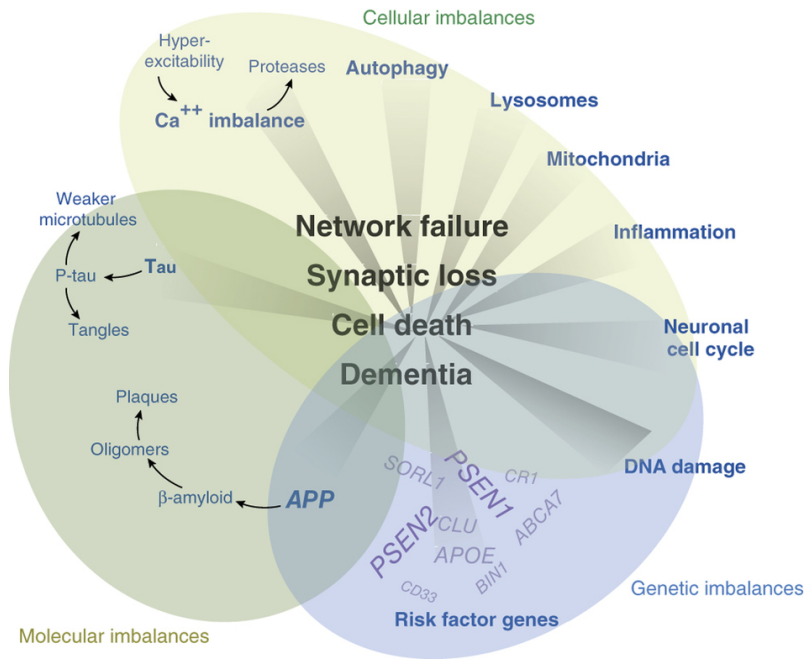
Increasing evidence suggests that AD pathology, particularly abnormal forms of tau, propagates in a prion-like manner [13], i.e. that Alzheimer's pathologies are transmitted from one neuron to another. However, the physical mechanism by which pathology is transported between neurons has not been identified [54]. The major alternative hypothesis is that of selective vulnerability, meaning that certain neurons are intrinsically more vulnerable to the pathogenic disease process than others [13, 54].

#### 1.1.5.4 Alternative mechanisms

A multitude of other mechanisms have been suggested to be involved in the pathogenesis of AD, such as immune responses and inflammatory, metabolic, and vascular factors. One suggestion is that there is a complex feedback loop between different mechanisms that causes AD [45] (Figure 1.6).

### 1.1.6 Biomarkers

The only direct measurement of AD pathology is through autopsy examination of the brain. Biomarkers are variables that can be measured in vivo to provide information about these pathological changes; these measurements are indirect indicators that the pathology exists. Disease biomarkers can be divided into a number of different groups. In the case of AD there is particular interest in those biomarkers that can assess the core pathologies of  $A\beta$ -plaque deposition and tau aggregation.



**Figure 1.6:** Hypothetical model of Alzheimer's disease as a complex interaction between cellular, molecular and genetic imbalances. Reprinted by permission from Macmillan Publishers Ltd: Nature Neuroscience [45], Copyright 2015.

Tau aggregation is not specific to AD, but is thought to mediate neurodegeneration. Whilst a stricter definition of the term biomarker is used here, which refers only to biological indicators of pathology, in the rest of this thesis a broader definition of the term is used, which encompasses any marker of the disease process. This includes measurements of higher level cognitive processes obtained from neuropsychological tests.

#### 1.1.6.1 $A\beta$ plaque deposition

The two most well validated biomarkers of brain  $A\beta$  plaque deposition are cerebrospinal fluid (CSF)  $A\beta_{1-42}$  and positron-emission tomography (PET)  $A\beta$  imaging. CSF  $A\beta_{1-42}$  ( $A\beta_{42}$ , Abeta,  $A\beta$ ) is a measure of the toxic amyloid- $\beta$  peptide implicated in AD. Low concentrations of CSF  $A\beta_{1-42}$  correlate with the clinical diagnosis of AD and  $A\beta$  pathology at autopsy [55–57]. A number of PET  $A\beta$  tracers are available, which allow the spatial pattern of amyloid deposition to be observed non invasively. The earliest tracer was Pittsburgh compound B (PiB), which binds to fibrillar  $A\beta$ , and shows a strong correspondence with  $A\beta$  deposition at

autopsy [58,59].

#### 1.1.6.2 Neurodegeneration

CSF total tau (t-tau,  $\tau$ -tau) and phosphorylated tau (p-tau,  $\rho$ -tau), fluorodeoxyglucose PET (FDG-PET) and structural magnetic resonance imaging (MRI) all provide different measures broadly implicated in neurodegeneration. Both CSF t-tau and p-tau increase in AD [60] and correlate well with neurofibrillary tau tangles at autopsy [61]. CSF t-tau concentration correlates with the severity of neurodegeneration, which is not specific to AD, whereas CSF p-tau reflects the formulation of AD-related neurofibrillary tangles [62]. FDG-PET is a measure of brain metabolism, in particular glucose uptake, and so can be used as an indication of the impaired synaptic activity associated with AD, with combined imaging and autopsy studies supporting this hypothesis [63]. The impairment of synapses is thought to begin before the observation of clinical symptoms but is also common to other neurological diseases. Measures of brain atrophy, a downstream consequence of neuronal cell loss, calculated from structural MRI scans also display a good correlation with neurodegeneration [64]. MRI based measures of atrophy (rate of volume loss) are not specific to AD but correlate well with Braak and Braak staging of tau pathology at autopsy [65,66], and can be used to determine the regional patterns of neurodegeneration characteristic to AD [67].

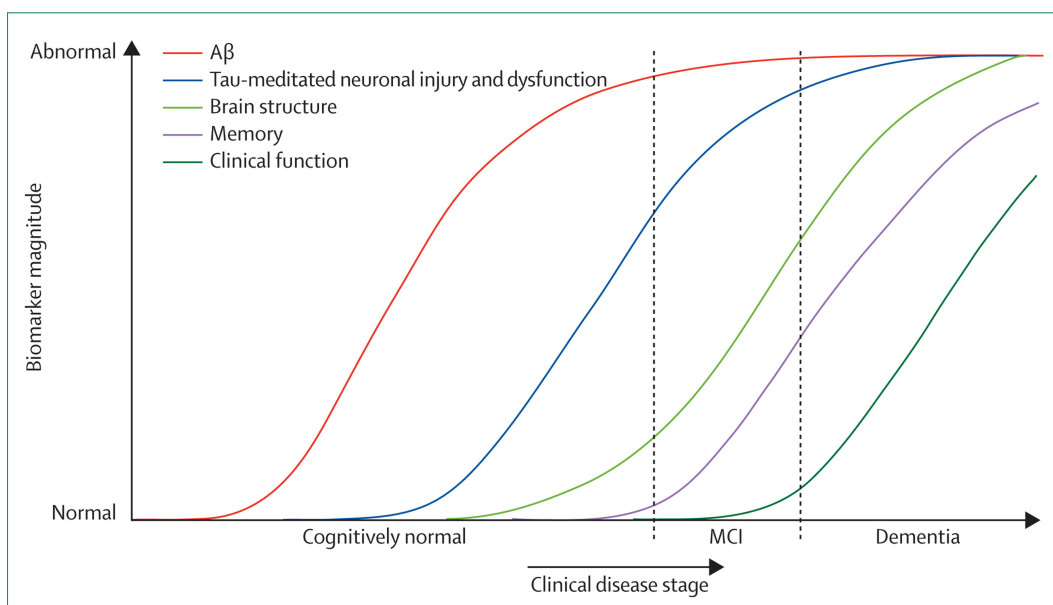
The recent development of tau tracers allows non invasive imaging of the pattern of neurofibrillary tangles [68–70]. These imaging techniques are undoubtedly exciting but require further validation of their reliability, quantitative performance and binding selectivity before they are widely used in research studies [71].

#### 1.1.6.3 Evolution of Alzheimer's disease biomarkers

Understanding the quantitative evolution of biomarkers in AD is of great interest for clinical trials in order to provide outcome measures and strategies for cohort selection. It is thought that clinical trials to date have taken place too late in the disease time course [72,73]. Biomarkers can potentially be used to identify and monitor the presymptomatic disease stages, during which treatments may be more effective. Additionally, a quantitative picture of biomarker progression can provide insights

into the underlying disease biology by, for example, indicating which is the initiating disease pathology, or elucidating interactions between different pathogenic processes.

Numerous hypothetical models of the progression of AD biomarkers have been proposed [74–79]. The most influential of these hypothetical models was published in 2010 by Jack et al. (Figure 1.7) [74], and has had a large impact within the neurology community. This model postulates that AD changes occur in the order:  $A\beta$ , tau-mediated neuronal injury and dysfunction, brain structure, memory, clinical function.  $A\beta$  is measured using CSF  $A\beta$  or amyloid PET; tau-mediated neuronal injury and dysfunction is measured using CSF p-tau, t-tau and FDG-PET; brain structure is measured using structural MRI; memory and cognition are measured using various cognitive tests.



**Figure 1.7:** Hypothetical model of Alzheimer's disease biomarker progression proposed by Jack et al. in 2010 [74]. Reprinted from [74], Copyright 2010, with permission from Elsevier.

Validation of hypothetical models is difficult because of the long disease time course, which is thought to span several decades [49, 80–82]. This means that it is challenging to follow subjects longitudinally, particularly during the pre-symptomatic phase, which can only be observed in large population-based observational studies. Whilst studies have provided support for some of the features

of hypothetical models [49, 80, 83, 84], they typically use crude staging measures based on clinical diagnoses or cognitive test scores. The result is a coarse picture of biomarker progression, which is unable to provide fine-grained disease staging information. Studies in dominantly-inherited AD [49] can provide a more fine-grained picture by identifying subjects who carry a pathogenic APP, PSEN1 or PSEN2 mutation presymptomatically, and indexing them by their estimated years to onset, which is based on parental age of onset. However, the applicability of these results to sporadic AD has yet to be determined.

### 1.1.7 Diagnostic criteria

#### 1.1.7.1 NINCDS-ADRDA criteria for probable Alzheimer's disease

A definitive diagnosis of AD requires microscopic examination of brain tissue at autopsy. The most widely used clinical diagnostic criteria for AD were established in 1984 by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) [85], commonly referred to as the NINCDS-ADRDA criteria. The criteria distinguish between probable, possible, and definite AD, where only a probable or possible diagnosis of AD can be made in vivo. The criteria for probable AD include: dementia established by clinical examination and confirmed by neuropsychological tests; deficits in two or more areas of cognition; progressive worsening of memory and other cognitive functions; onset between ages 40 and 90, most often after age 65; additional criteria that rule out other conditions.

#### 1.1.7.2 Updated probable Alzheimer's disease criteria

More recent diagnostic criteria for probable AD include the use of biomarker evidence of AD as a supportive feature [86–89]. The criteria proposed by McKhann et al. in 2011 [88], for example, subdivide the diagnosis of probable AD into probable AD with and without evidence of the AD pathophysiological process, where the core criteria for probable AD are similar to the NINCDS-ADRDA criteria. Evidence of the AD pathophysiological process is divided into biomarker evidence of  $A\beta$  from PET or CSF, and biomarker evidence of neuronal injury from CSF tau,

FDG-PET or structural MRI, with evidence of both  $A\beta$  and neuronal injury providing stronger support for the diagnosis.

### 1.1.7.3 Mild cognitive impairment

A more subtle diagnostic entity, mild cognitive impairment (MCI), has also been proposed [90]. Although this label is widely used in research studies, the definition and subcategorisation of the term varies. Criteria for MCI typically consist of subjective memory or cognitive symptoms, objective memory or cognitive impairment, and unaffected activities of daily living [86]. Often a distinction is made between amnestic and non-amnestic MCI, depending on the presence of significant memory complaints. Amnestic MCI is intended to be more specific to AD, however, clinical follow-up and neuropathological studies show that all of the aforementioned diagnoses lack specificity for AD [91,92]. In 2011 Albert et al. [93] proposed the use of two sets of criteria for diagnosing MCI: core clinical criteria that can be used without access to biomarker information, and research criteria for use in research settings and clinical trials. The criteria have four levels of confidence depending on the presence and strength of biomarker findings. The core clinical and cognitive criteria include cognitive concerns and objective evidence of cognitive impairment, together with preserved activities of daily living and severity not reaching the threshold for a dementia diagnosis. If biomarker measurements are available, three more specific diagnoses can be made: (A) MCI due to AD with intermediate likelihood, (B) MCI due to AD with high likelihood, (C) MCI unlikely due to AD. Positive biomarker evidence of either  $A\beta$  or neuronal injury supports (A), positive biomarker evidence of both  $A\beta$  and neuronal injury supports (B), and negative biomarker evidence of both  $A\beta$  and neuronal injury supports (C). Biomarker evidence of  $A\beta$  is measured using CSF or PET; biomarker evidence of neuronal injury is measured using CSF tau, FDG-PET or structural MRI.

### 1.1.7.4 Preclinical Alzheimer's disease

In recent years the development of presymptomatic diagnostic criteria has been advocated following the accumulation of evidence suggesting that biomarker changes in AD begin decades before symptom onset [49, 80–82]. Presymptomatic diagnos-



tic criteria therefore focus on identifying the AD pathophysiological process rather than the AD dementia syndrome. In 2011 Sperling et al. [94] proposed preclinical criteria consisting of three stages: (1) asymptomatic cerebral amyloidosis, with evidence of  $A\beta$  from PET or CSF, (2) asymptomatic amyloidosis and downstream neurodegeneration, with additional evidence of markers of neuronal injury from CSF tau, FDG-PET or structural MRI, (3) amyloidosis, neuronal injury and subtle cognitive/behavioural decline, with additional evidence of subtle cognitive change. These preclinical criteria are not well established due to their dependence on hypothetical models of the evolution of biomarker measurements in AD.

#### 1.1.7.5 Neuropathologic Alzheimer's disease

The growing consensus on the need to distinguish the AD pathophysiological process from the AD dementia syndrome has led to the revision of the neuropathologic criteria for assessing AD at autopsy. The guidelines proposed by Hyman et al. in 2012 [19] remove the requirement of a history of dementia from the criteria, instead focussing on the evaluation of three different neuropathologic parameters: (A)  $A\beta$  plaque score (evaluated using Thal et al. 2002 [23]), (B) neurofibrillary tangle stage (evaluated using Braak and Braak 1991 [26] or Braak et al. 2006 [95]), (C) neuritic plaque score (evaluated using CERAD proposed by Mirra et al. 1991 [96]). The presence of both  $A\beta$  plaques and neurofibrillary tangles is necessary for the identification of AD neuropathologic change. Neuritic plaques are a particular type of  $A\beta$  plaque that consist of a dense  $A\beta$  protein core surrounded by dystrophic neurites. This form of  $A\beta$  plaque is the most closely associated with neuronal injury. The neuropathologic criteria proposed by Hyman et al. also emphasise the importance of assessing non-AD brain lesions to identify the presence of comorbid conditions.

#### 1.1.8 Treatments

There are currently five FDA-approved drugs for AD, which mask the symptoms of the disease, but do not modify the underlying disease process. They fall into two categories: cholinesterase inhibitors and memantine [97]. Cholinesterase inhibitors slow down the process that breaks down the neurotransmitter acetylcholine; me-

mantine regulates the activity of glutamate, a neurotransmitter involved in learning and memory. The major focus of drug companies is developing disease modifying treatments, with promising targets including  $A\beta$ ,  $\beta$ -secretase,  $\gamma$ -secretase, tau protein, inflammation, and insulin resistance [97]. The results of clinical trials to date have been disappointing, which is thought to be due to treatments being administered too late in the disease, at which point irreversible pathogenic processes have already taken place [72, 73].

## 1.2 Research problem

### 1.2.1 Problem statement

The temporal progression of AD is not well understood. There are two key aspects to this problem.

1. The underlying biological mechanisms that cause AD have not been determined.
2. The quantitative evolution of biomarkers in AD has yet to be characterised.

The work presented in this thesis focusses on problem 2.

### 1.2.2 Justification

Despite 110 years of Alzheimer's research and an exponential growth in the amount of research since 1975 [11], still relatively little is understood about the causes and progression of AD.

Although a vast number of molecular, cellular and genetic imbalances have been implicated in AD, how these imbalances give rise to AD is still unknown (Problem 1, Section 1.2.1) [45]. In particular, the precise mechanism by which the two major proteins involved in AD,  $A\beta$  and tau, produce downstream neurodegeneration and cognitive deficits remains unclear [13], as does the mechanism by which they spread to different brain regions [13, 54].

At the biomarker level, the quantitative evolution of AD has yet to be determined (Problem 2, Section 1.2.1). Hypothetical models have been proposed that

describe how biomarkers evolve with disease progression [74–79], but these models are not quantitative and their validation to date has typically been at a coarse scale [83, 84].

### 1.2.3 Motivation

Better understanding of the causes and progression of AD will facilitate drug development. A mechanistic understanding of AD (Problem 1, Section 1.2.1) would identify the key targets for drug development, and the appropriate window in the disease for intervention. A quantitative model of biomarker evolution (Problem 2, Section 1.2.1) is essential for precision medicine. Such a model could enable patient stratification into targeted groups for particular therapies, and provide outcome measures for clinical trials, which can monitor the effectiveness of different treatments. If a disease-modifying treatment is discovered, quantitative biomarker models could be used in the clinic to provide disease staging and prognostic information, and to assign patients to different treatments. Besides enabling personalised medicine, quantitative models of biomarker evolution can provide insights into the underlying disease mechanisms by indicating which are the earliest pathologies, and by elucidating interactions between different pathogenic processes.

## 1.3 Thesis contributions

This thesis investigates the progression and heterogeneity of AD by developing mathematical models that can characterise the temporal evolution of disease biomarkers. The models only require cross-sectional or short-term longitudinal observations to reconstruct the full temporal progression of the disease, meaning that they can be fit to widely available datasets such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset. Three key contributions are made, as follows.

### 1.3.1 Data-driven models of biomarker changes in sporadic and dominantly-inherited Alzheimer’s disease

The first contribution of this thesis is the development of data-driven models of disease biomarker changes for application to sporadic and dominantly-inherited

AD. These models extend the event-based model (EBM) developed by Fonteijn et al. [98] to allow for heterogeneity in the control and disease population, and missing biomarker values. The EBM describes disease progression as a series of events, where each event corresponds to a new biomarker becoming abnormal. Application of these models to sporadic and dominantly-inherited AD reveals the sequence in which biomarkers become abnormal in each form of AD. This thesis further explores the utility of the models for patient staging. Specifically, the following sub-contributions are made.

- (a) Adaptation of the EBM in [98] to allow for heterogeneity in the control and disease populations. Without this adaptation the EBM is not applicable to sporadic AD due to its reliance on a well-defined control population.
- (b) Application of (a) to sporadic AD to determine the sequence in which biomarkers become abnormal, and to demonstrate the utility of the model for patient staging. Previous attempts to determine biomarker ordering were dependent on the use of clinical diagnoses to stage patients, or on the use of thresholds to define biomarker abnormality.
- (c) Adaptation of the EBM in [98] to allow for missing data. Prior to this adaptation the EBM was not applicable for modelling the evolution of multimodal biomarker data collected in dominantly-inherited AD datasets due to small numbers of subjects with data spanning multiple biomarkers.
- (d) Application of (c) to dominantly-inherited AD to determine the sequence of biomarker abnormality in dominantly-inherited AD. Previous studies estimating biomarker ordering in dominantly-inherited AD relied on the use of estimated years to onset (based on parental age of onset) to stage participants.

### **1.3.2 A simulation system for biomarker evolution in neurodegenerative disease**

The second contribution of this thesis is the development of a simulation framework for biomarker evolution in neurodegenerative disease, facilitating the evaluation of

the performance of data-driven disease progression models. The following sub-contributions are made.

- (a) Development of a simulation framework that can generate synthetic neurodegenerative disease datasets for evaluating the performance of data-driven disease progression models. In particular, the simulation system can synthesise heterogeneous neurodegenerative disease populations, including disease subtypes that have distinct patterns of biomarker evolution.
- (b) Application of the simulation framework to perform a stability analysis of the EBM developed in Contribution 1.3.1 to ascertain the sensitivity of the EBM to different types of heterogeneity. A stability analysis of another data-driven disease progression model - a differential equation model (DEM) - is performed to compare the robustness of the two data-driven models.

### **1.3.3 A data-driven model of disease subtypes with distinct patterns of biomarker evolution in frontotemporal dementia and Alzheimer's disease**

The final contribution of this thesis is the development of a data-driven model of disease subtypes with distinct patterns of biomarker evolution. This model describes disease progression as groups of individuals who have a common sequence of events, where each event corresponds to a biomarker reaching a particular z-score compared to controls. This model is validated using data from genetic frontotemporal dementia (FTD), which has distinct genetic subtypes. Application of this model to sporadic AD reveals data-driven subtypes of AD that have distinct sequences in which biomarker measurements reach different severity levels (modelled as z-scores relative to a control population). The utility of this subtyping model for patient staging is further demonstrated. More specifically, the following sub-contributions are made.

- (a) Development of a dynamic clustering technique that allows population subgroups with distinct sequences of biomarker changes to be recovered. In this

model biomarker changes are indexed as z-score events, such that each z-score event corresponds to a biomarker reaching a particular z-score relative to a control population. Previous modelling techniques depended on the assumption that all subjects follow a single common pattern of biomarker evolution, or that all subjects are at a single disease stage.

- (b) Validation of the dynamic clustering technique using data from genetic FTD. The sequence of biomarker changes for each genetic subgroup is determined by fitting a single dynamic cluster to each of the genetic subgroups: GRN, MAPT and C9orf72 mutation carriers. The ability of the dynamic clustering model to recover the sequence of each subgroup without prior knowledge of the group labels is demonstrated.
- (c) Application of dynamic clustering to sporadic AD to determine subtypes of AD with distinct sequences of biomarker changes, and to establish the utility of these subtypes for patient stratification. Prior studies that clustered AD biomarker measurements depended on the assumption that subjects were at a single disease stage, meaning that the temporal evolution of each subtype could not be determined.

## 1.4 Structure of this thesis

This chapter (**Chapter 1**) has provided background information, motivation and context for the research problem and contributions of this thesis. The rest of this thesis is structured as follows.

**Chapter 2** reviews the state of the art in AD progression modelling. This chapter critically assesses the range of different progression models that have been applied to AD to establish which models are the most promising for evaluating the quantitative evolution of AD biomarkers, and to identify the limitations of current modelling techniques.

**Chapter 3** presents sub-contributions (a) and (b) of contribution 1.3.1. I propose an adaptation of the EBM [98] for use with multi-modal sporadic disease datasets. I apply this model to determine the sequence of biomarker abnormality

in sporadic AD, and further demonstrate the models utility as a patient staging system.

**Chapter 4** details sub-contributions (c) and (d) of contribution 1.3.1. I develop an adaptation of the EBM [98] for use with missing data. I apply this model to determine the order in which biomarkers become abnormal in dominantly-inherited AD without reliance on the use of estimated years to onset to stage participants.

**Chapter 5** formulates a simulation framework for biomarker evolution in neurodegenerative disease (contribution 1.3.2). I use this framework to perform a stability analysis of the EBM and a DEM of disease progression by fitting the models to synthetic datasets with different types of heterogeneity.

**Chapter 6** presents an initial formulation of the dynamic clustering technique developed in Chapter 7.

**Chapter 7** builds on the work in Chapter 6 to develop a more parsimonious dynamic clustering model (contribution 1.3.3) that can elucidate population subgroups with distinct patterns of biomarker evolution, enabling the identification of data-driven disease subtypes. This technique is validated by demonstrating the ability to recover the progression patterns of known genetic subgroups in FTD. Dynamic clustering is then applied to sporadic AD data to reveal different subtypes of AD and the sequence in which biomarkers reach various z-scores for each subtype. The utility of the model for patient staging and stratification is further demonstrated.

**Chapter 8** discusses opportunities for further work arising from this thesis, as well as interesting future directions for data-driven disease progression modelling in general.

**Chapter 9** summarises the work presented in this thesis.





## Chapter 2

# State of the art in Alzheimer's disease progression modelling

This chapter reviews the state of the art in AD progression modelling to provide context for the contributions of this thesis (Section 1.3, Chapter 1). Section 2.7 is based on a review article I wrote for *Advances in Clinical Neuroscience and Rehabilitation*, in collaboration with Neil Oxtoby, Jonathan Schott and Daniel Alexander.

## 2.1 Associated publications

A. L. Young, N. P. Oxtoby, J. M. Schott, and D. C. Alexander. Data-driven models of neurodegenerative disease. *Advances in Clinical Neuroscience and Rehabilitation*, 14(5):6–9, 2014

## 2.2 Introduction

This thesis addresses the problem of characterising the temporal progression of AD by developing mathematical models that quantify the evolution of disease biomarkers. This review examines the range of different progression models that have been applied to AD. In this review I first summarise the literature on two types of non-biomarker model: neuropathological models and animal models, which provide interesting qualitative biological insights but do not directly provide quantitative measures of disease stage. I then review various biomarker models to establish what is known about the quantitative evolution of AD biomarkers. I first review

scalar biomarker models that use single dimension biomarkers and traditional statistical analysis techniques, which require subjects to be indexed by disease stage. I then review high dimensional biomarker models, which learn new biomarkers from high dimensional data but still require subjects to be indexed by disease stage. Finally, I review the literature on data-driven biomarker models, which develop novel statistical methodology to allow the temporal progression of biomarkers to be reconstructed without prior knowledge of each individual's stage along the disease time course. These data-driven models provide the basis for the novel techniques developed in this thesis.

## **2.3 Neuropathological models**

The earliest models of AD progression are those derived from neuropathological studies (e.g. [23, 26, 96]). Models of neurofibrillary tangle accumulation and amyloid deposition have been proposed based on the cross-sectional distribution of pathology at autopsy. These models typically depend on the assumption that there is a common pattern of pathology for all subjects, with each subject representing a particular stage along this single progression pattern.

### **2.3.1 Neurofibrillary tangle accumulation**

The most influential model of neurofibrillary tangle accumulation is that of Braak and Braak in 1991 (Figure 1.3) [26], which is still used to diagnose pathological AD at autopsy [19]. Braak and Braak proposed a model that consists of six stages based on the distribution of neurofibrillary tangles and neuropil threads: stages I-II are confined to the transentorhinal region, stages III-IV involve both the entorhinal and transentorhinal layer, stages V-VI consist of isocortical destruction. Braak and Braak also proposed three stages of amyloid plaque deposition, but they chose to index their model by neurofibrillary tangle stage as they observed that the distribution of amyloid plaques varied widely between architectonic units and individuals. In 2006 Braak et al. updated the model to incorporate methodological developments in neuropathology [95]. This model was subsequently incorporated into neuropathological AD diagnostic criteria [19].

The Braak model [26, 95] assumes that there is a common pattern of neurofibrillary tangle deposition across subjects. More recently, Murray et al. [27] found that approximately 25% of AD cases do not follow the stereotypical pattern of neurofibrillary tangle deposition described by the Braak model. By considering cases with a Braak stage of greater than IV, and constructing an algorithm that classifies AD cases into ‘typical’, ‘hippocampal sparing’ or ‘limbic predominant’, they found that 11% of the 889 AD cases they analysed were hippocampal sparing, and 14% of cases were limbic predominant. This percentage is somewhat artificial as their algorithm is based on percentile cut-offs for the ratio of hippocampal to cortical neurofibrillary tangle count, but this study does clearly indicate the presence of atypical AD pathology. This heterogeneity may represent a pathological spectrum ranging from hippocampal sparing to limbic predominant, or three distinct subgroups. Another interesting finding of this study was that clinical presentation, age at onset, disease duration and rate of cognitive decline differed between the different AD subtypes. Hippocampal sparing cases more commonly had an atypical clinical diagnosis, a younger age of onset, a shorter disease duration and a faster rate of cognitive decline than typical and limbic predominant AD cases.

### 2.3.2 Amyloid plaque deposition

There are two widely used models of amyloid plaque deposition for neuropathological AD diagnosis [19]: the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) criteria proposed by Mirra et al. in 1991 [96] and the model proposed by Thal et al. in 2002 [23].

The CERAD criteria are coarse semi-quantitative criteria based on the assessment of neuritic amyloid plaques in the neocortex. The procedure involves taking samples of brain tissue from three areas of the neocortex: the superior and middle temporal gyri, the middle frontal gyrus, and the inferior parietal lobule, and rating them as having either ‘sparse’, ‘moderate’ or ‘frequent’ neuritic amyloid plaques. The overall plaque density is the maximum plaque density across the three regions of the neocortex. The age-related CERAD plaque score: 0, A, B or C, where C indicates the most confident evidence of neuropathological AD, is derived by comparing

the overall plaque density with the expected plaque density for that age group. The CERAD criteria are used as a measure of dense-core neuritic plaques in modern pathological AD diagnostic criteria [19].

The model proposed by Thal et al. in 2002 (Figure 1.2) [23] has five stages and is based on the spatial distribution of  $A\beta$  deposits, which include diffuse as well as neuritic plaques. Phase 1 of their model consists of  $A\beta$  deposits in any region of the neocortex. In the subsequent phases there are additional  $A\beta$  deposits in allocortical regions (phase 2), diencephalic nuclei and the striatum (phase 3), some brainstem nuclei (phase 4), and finally the cerebellum and other brainstem nuclei (phase 5).

### **2.3.3 Critical assessment**

Neuropathological models provide interesting qualitative insights into the spatial progression of AD pathology and are essential for the neuropathological confirmation and staging of AD. In particular, they show that the distribution of neurofibrillary tangles is more stereotypical across subjects than the distribution of amyloid plaques. Moreover, they demonstrate that whilst the overall amyloid plaque and neurofibrillary tangle burden are correlated, their spatial progression patterns are markedly different, and the relative amount of amyloid plaque and neurofibrillary tangle pathology varies across individuals.

The utility of neuropathological models for in vivo patient staging is limited by several factors. First, the models are ex vivo. For application in vivo biomarkers would have to be used, which are correlated with, but do not measure exactly the same pathology. Second, the models are not quantitative, which limits their accuracy. Third, the time scale and spatial resolution of the models is coarse, with typically six or fewer stages and large brain regions. Finally, each model only measures a single pathology, so it is not straight forward to integrate the stages from the two different pathologies.

## 2.4 Animal models

Genetically modified mouse models<sup>1</sup> have provided insights into the underlying mechanisms of AD [100].

### 2.4.1 Transgenic mouse models

In transgenic mouse models a foreign gene is introduced, which can be from another species. In transgenic mice that overproduce mutant APP, A $\beta$  accumulates into extracellular plaques. This accumulation is accelerated when A $\beta$ <sub>42</sub> is preferentially cleaved from APP compared to A $\beta$ <sub>40</sub>, and leads to earlier and more severe cognitive decline [100]. Most transgenic Alzheimer's mouse models show cognitive deficits before extracellular A $\beta$  plaques accumulate. This observation led to the hypothesis that soluble A $\beta$  oligomers may be the pathogenic mechanism in AD [100]. Interestingly, APP-overexpressing mice do not develop neurofibrillary tangles, but do show hyperphosphorylation of tau. Additional gene alterations, such as mutated human tau, are necessary to induce neurofibrillary tangles in mice [100].

### 2.4.2 Knockout mouse models

In knockout mouse models a particular gene is removed. These mice do not model the disease process itself, but are instead used to elucidate the function of specific genes. Presenilin 1 knockout mice suggest that presenilin and the  $\gamma$ -secretase complex might have additional functions that are not related to the production of A $\beta$ , with PS1 knockout mice having developmental defects in the central nervous and skeletal systems [100]. Another interesting finding is that when APP transgenic mice are crossed with tau knockout mice, the cognitive deficits associated with APP and A $\beta$  are prevented [101], even though the absence of tau does not affect the development of A $\beta$  plaques. This suggests that tau is necessary to mediate the pathogenic effects of A $\beta$  on cognition.

---

<sup>1</sup>An interactive summary of the findings from different mouse models of AD can be found at [www.alzforum.org/research-models](http://www.alzforum.org/research-models).

### 2.4.3 Critical assessment

There are several concerns regarding the translatability of animal models [100]. First, the models are based on dominantly-inherited AD, rather than sporadic AD. Second, the synapse and cell loss in the models is not substantial, so they may only be representative of the prodromal phase of AD. Third, there is discordance between the results of preclinical animal models and human clinical trials.

## 2.5 Scalar biomarker models

The simplest AD biomarker models are those that use scalar biomarkers and traditional statistical analysis techniques, which require knowledge of a subjects disease stage in order to position them along the time axis.

### 2.5.1 Comparing diagnostic groups

There have been a large number of studies comparing biomarkers between diagnostic groups (e.g. [102–106]). These studies provide a coarse picture of which biomarkers are dynamic at each disease stage, with evidence suggesting that CSF  $A\beta$  and PiB-PET become abnormal while subjects are still cognitively normal (CN), and MRI, FDG-PET and CSF tau are already abnormal in those with mild cognitive impairment (MCI) [74]. However, the use of clinical diagnoses to stage patients limits the temporal resolution of these studies, typically to just three stages: CN, MCI and AD. Iturria-Medina et al. [107] perform a more complex statistical analysis of the expected biomarker trajectories for subjects at different disease stages, which takes into account age, gender, education and number of APOE  $\epsilon 4$  alleles. Their results broadly agree with other studies, but suggest that vascular dysregulation may also play an early role in AD. However, the temporal resolution of their model is still fundamentally limited by the accuracy of the clinical diagnoses.

### 2.5.2 Using cut points to define biomarker abnormality

Jack et al. [83] have proposed that the sequence in which biomarkers become abnormal can be inferred by observing the proportion of subjects within each diagnostic group that have abnormal biomarker levels. They used this technique to validate

the ordering of a subset of the biomarkers (CSF  $A\beta_{1-42}$ , CSF tau and hippocampal volume) proposed by their hypothetical model a year previously (Figure 1.7) [74]. They concluded that CSF  $A\beta_{1-42}$  becomes abnormal first, then CSF tau, and then hippocampal volume. However, their results are dependent on choosing cut points defining abnormal biomarker levels, which are not easy to establish [108], and are likely to affect the sequence of biomarker abnormality predicted by their model. Moreover, their method does not easily extend to larger numbers of biomarkers.

### 2.5.3 Indexing by cognitive test scores

An obvious approach to reconstructing fine-grained biomarker trajectories is to index subjects by their cognitive test scores rather than their clinical diagnoses. Caroli et al. [109] used such an approach to investigate the dynamics of four AD biomarkers: CSF  $A\beta_{1-42}$ , CSF tau, hippocampal volume and FDG-PET, as a function of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-cog). They found that CSF  $A\beta_{1-42}$  was the first to become abnormal and stabilises early on in the disease time course, whereas CSF tau, hippocampal volume and FDG-PET become abnormal later in the disease time course and continue to change as the disease progresses. Sabuncu et al. [110] performed a similar analysis of hippocampal volume loss and cortical thinning as a function of Mini Mental State Exam (MMSE) score. They found that hippocampal volume loss and cortical thinning both accelerate early on in the disease and have a sigmoidal shape. However, there are two major limitations to using cognitive test scores as a measure of disease progression. First, cognitive test scores are known to have floor and ceiling effects, in particular they can't be used to measure the presymptomatic disease stages. Second, the accuracy of the recovered progression patterns is limited by the accuracy of the cognitive test scores, which are imprecise and often have practice effects.

### 2.5.4 Indexing by estimated years to onset

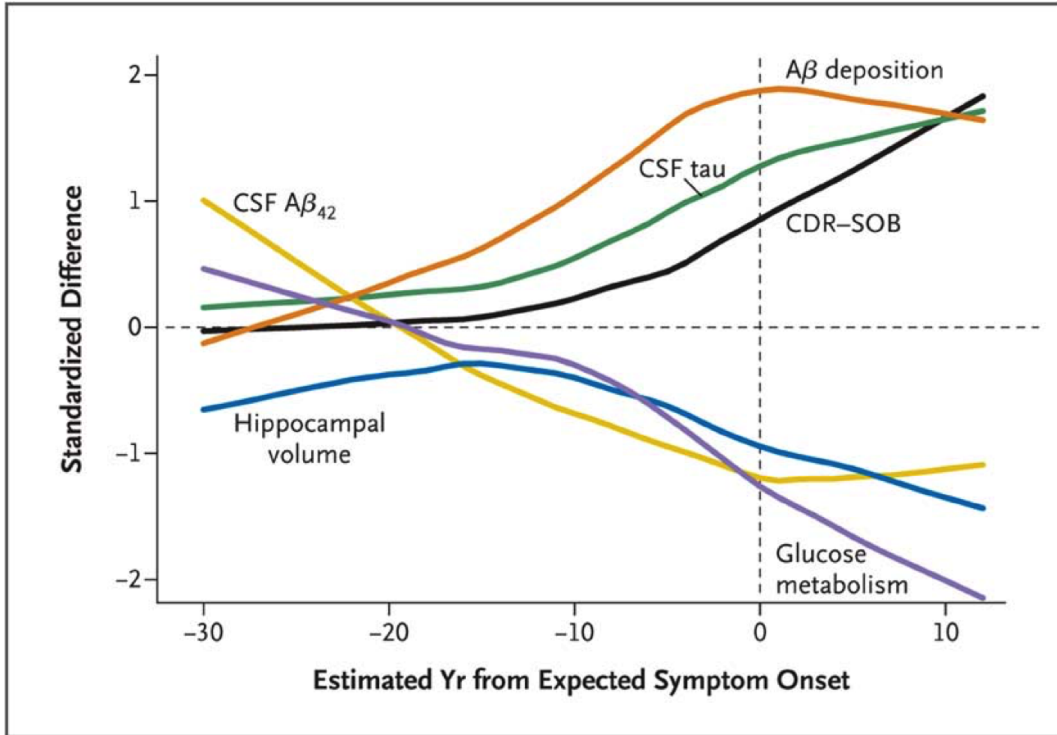
In dominantly-inherited AD it is possible to identify subjects presymptomatically by genetically screening families known to carry a pathogenic mutation. In those that carry the genetic mutation, it is also possible to estimate their time to disease

onset based on their parent's age of onset. In 2012 Bateman et al. (Figure 2.1) [49] constructed a model of biomarker changes in dominantly-inherited AD from 128 participants in the Dominantly Inherited Alzheimer Network (DIAN) study. By regressing participants biomarker values against their estimated years to onset, they found that: CSF  $A\beta_{42}$  levels became abnormal 25 years before expected onset;  $A\beta$  deposition (measured on PiB-PET), CSF tau and hippocampal volume became abnormal 15 years before expected onset; cerebral hypometabolism (measured on FDG-PET) and episodic memory (measured using story A from the Logical Memory subtest of the Wechsler Memory Scale-Revised) became abnormal 10 years before expected onset; global cognitive impairment (measured using the MMSE and the Clinical Dementia Rating scale) became abnormal 5 years before expected onset; participants met the criteria for an AD diagnosis 3 years after expected onset. However, the applicability of these results to the more common sporadic form of AD remains to be determined. An additional concern is the validity of estimating the participants time to onset from their parent's age of onset.

### **2.5.5 Indexing by conversion between diagnoses**

Several studies have constructed population-level biomarker trajectories by retrospectively indexing subjects according to their time to conversion between diagnostic categories. This approach requires that subjects are observed until conversion. Buchhave et al. [80] performed such an analysis of CSF measures in 137 subjects with MCI who converted to AD. Their study had a median follow-up time of 9.2 years (range 4.1 to 11.8 years). By indexing subjects according to whether they were between 0 and 2.5, 2.5 and 5, or 5 and 10 years from conversion they found that CSF measures of  $A\beta_{1-42}$  were already abnormal between 5 and 10 years before disease onset. They found that CSF p-tau and t-tau changes also became abnormal early, but that the changes were more gradual. More advanced statistical models have also been proposed by Schmidt-Richberg et al. and Guerrero et al. based on the idea of indexing subjects by time to conversion [111, 112]. These models allow probabilistic estimation of a new patient's disease progress and prediction of their time to conversion between diagnoses. However, all of the aforementioned models





**Figure 2.1:** Biomarker changes in dominantly-inherited Alzheimer’s disease estimated by Bateman et al. in 2012 [49] by regressing against expected years to onset. Reproduced with permission from [49], Copyright Massachusetts Medical Society.

are limited by the accuracy of the clinical diagnoses used to stage patients. Moreover, the requirement that a large population are observed until conversion limits the datasets that these models can be fitted to, as well as their ability to model population heterogeneity. The long follow-up requirement further prevents these models from being applied to new biomarkers.

### 2.5.6 Critical assessment

Scalar biomarker models are highly relevant clinically as they can be used to amalgamate information from well-established disease biomarkers in order to provide patient staging and prognostic systems. Models to date suggest that CSF  $A\beta$  and PiB-PET become abnormal early in AD, followed by CSF tau, FDG-PET and hippocampal volume, and then cognitive test scores.

The temporal resolution of current models is limited by their dependence on clinical staging information to determine an individual’s position along the disease

time course. Positioning subjects based on their clinical diagnosis limits the accuracy of the recovered progression pattern to just three stages: CN, MCI and AD. Alternative techniques for modelling scalar biomarker trajectories either require the use of cut points to define abnormal biomarker levels, are based on cognitive test scores and so cannot track the long prodromal period of AD, are only applicable to dominantly-inherited AD, or require large retrospective studies of subjects who convert between clinical diagnoses.

## **2.6 High dimensional biomarker models**

More complex AD biomarker models consider high-dimensional data, such as image data, allowing more subtle patterns of disease progression to be uncovered. However, these models still use statistical analysis techniques that require knowledge of a subject's disease stage in order to position them along the time axis.

### **2.6.1 Comparing diagnostic groups**

Numerous studies have looked at voxelwise imaging differences, or other derived measures such as shape changes, between different diagnostic groups and controls [67, 113–120]. Structural MRI studies (e.g. [67, 113–116]) have shown that the pattern of atrophy mirrors the sequence of neurofibrillary tangle deposition estimated by neuropathological models such as the Braak and Braak model [26], with the entorhinal cortex and hippocampus being the earliest structures to atrophy. In PiB-PET studies, PiB retention has been observed in the frontal cortex, the precuneus and posterior cingulate, the temporal and parietal cortices, the occipital cortex and lateral temporal cortex, and the striatum [118, 119]. In FDG-PET studies, it has been found that cerebral metabolism is reduced, particularly in the temporal and parietal cortices. The spatial resolution of PiB-PET and FDG-PET is not as high as structural MRI and so patterns are frequently reported as an overall level of amyloid deposition or glucose metabolism [118, 119]. Diffusion imaging can also be used to look at white matter microstructure in AD, but is less well established [120]. Whilst the aforementioned studies are able to provide detailed spatial pictures of AD progression, their temporal resolution is limited by the requirement that subjects are

indexed by disease stage.

### **2.6.2 Prediction of conversion between diagnoses**

A multitude of studies have looked at predicting conversion between different diagnostic categories using images from various modalities [117, 121] or by combining different biomarkers [122]. These studies can provide valuable insights into which regions are the most discriminative for predicting conversion between two diagnoses at a particular snapshot in the disease time course:  $X$  years to conversion. However, they are typically limited to modelling a single disease stage.

### **2.6.3 Network models**

Network models describe how a disease propagates along structural or functional brain networks [123–127]. The data they require are high dimensional regional connectivity patterns, but this dimensionality is reduced by using a mechanistic approach, which physically constrains the solution, or by considering major components of the network, for example by using independent component analysis to extract intrinsic connectivity networks. In 2009, Seeley et al. [123] performed a study that compared intrinsic functional and structural networks in healthy subjects with the atrophy patterns of five neurodegenerative diseases. They found that the intrinsic functional and structural networks were correlated with the different atrophy patterns. Their results suggest that each structural and functional brain network has a selective vulnerability to a particular neurodegenerative disease. Other network models [124, 125, 127] have supported the transneuronal spread hypothesis (i.e. the prion hypothesis), whereby misfolded proteins physically spread from one neuron to another along brain networks. One approach [125, 127] is to use mechanistic network diffusion models and compare predicted atrophy patterns from a particular seed region with actual atrophy patterns. An alternative approach [124] is to derive hypotheses about the expected atrophy patterns under different network spreading mechanisms and compare the predicted patterns to observed atrophy patterns. Network models provide valuable mechanistic insights into AD, which are important for drug development, but they are not directly applicable for disease staging and

monitoring. Raj et al. [126] have provided some evidence that network diffusion models can be used to predict future atrophy and hypometabolism patterns, but the clinical utility of these predicted patterns has yet to be established. Additionally, the validity of network models is based on their alignment with a single coarse end stage disease atrophy pattern, and the use of a group-level functional or structural network connectivity pattern from healthy subjects.

#### **2.6.4 Critical assessment**

High dimensional biomarker models can provide valuable new insights into disease progression patterns by finding complex relationships within or between different data types. They are less directly clinically applicable because novel biomarkers require further validation before being widely used. Models to date have suggested that the pattern of regional atrophy mirrors the sequence of neurofibrillary tangle deposition, and that misfolded proteins may spread in a prion-like manner from one neuron to another.

As with the scalar biomarker models, the temporal resolution of current high-dimensional biomarker models is limited by their dependence on clinical staging information to determine an individual's position along the disease time course.

### **2.7 Data-driven biomarker models**

Traditional statistical analysis techniques estimate biomarker trajectories by assuming a priori knowledge of where each data point lies along the disease time course. Hence, the majority of studies of neurodegenerative disease biomarker progression (e.g. [67, 114]) rely on the use of a priori clinical classification as a patient staging measure and then compare biomarkers across groups. This reliance on clinical staging limits the temporal resolution of the biomarker progression to only a few stages, e.g. in AD there are typically just three stages: CN, MCI and AD. Recently a new family of truly data-driven statistical models (e.g. [81, 98, 128]) have emerged that do not require prior knowledge of the stage of each individual along the disease time course. This is a major advantage, as it allows for a complete picture of disease progression incorporating the full set of biomarkers, and with much higher

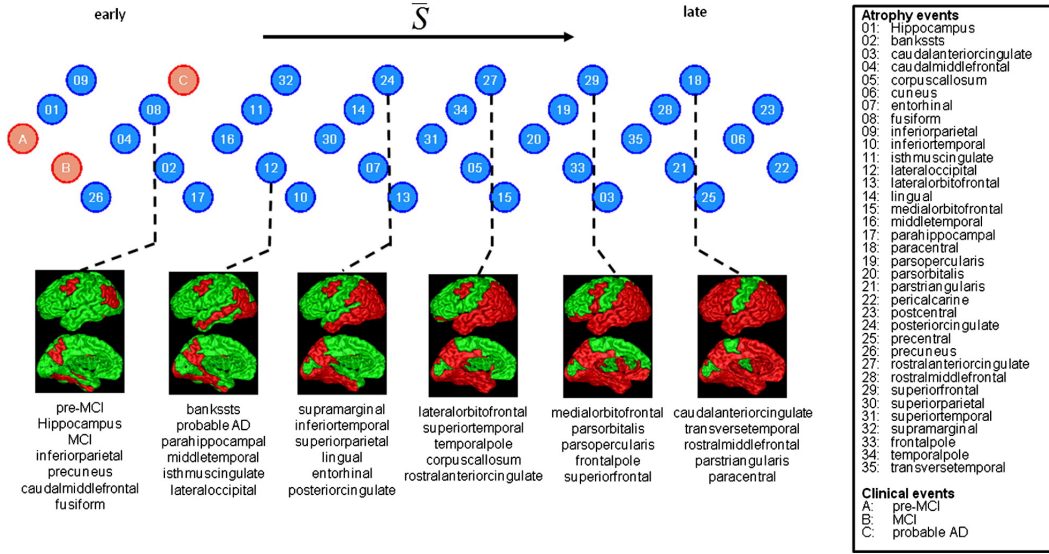
temporal resolution.

### 2.7.1 The event-based model

The event-based model (EBM) (Figure 2.2) [98] describes disease progression as a series of events, where each event corresponds to a particular biomarker becoming abnormal. The unique property of the EBM is that it directly encodes, and thus estimates from the data, the ordering in which biomarkers become abnormal, or, more strictly, observably different from normal levels. This sequence of events provides a simple and intuitive description of disease progression, as well as a natural patient staging system - at stage X, the first X events have occurred. The EBM has been applied to recover the sequence of regional neurodegeneration in both dominantly-inherited AD and Huntington's disease [98]. Another key strength of the EBM is its probabilistic formulation, which provides measures of confidence in both the sequence of biomarker abnormality events across the population, and an individual's model stage. The EBM naturally extends to differential diagnosis by providing a likelihood of each candidate neurodegenerative disease, which is achieved by fitting an individual's set of biomarker measurements to each corresponding biomarker sequence. One limitation of the EBM is that it doesn't incorporate information on the time between events or the rate of biomarker decline, which somewhat limits its utility for prognosis and monitoring. Another limitation is its reliance on a well-defined control population, which means the EBM is not currently applicable to sporadic AD.

### 2.7.2 Differential equation models

Differential equation models (DEMs) [81, 82, 110, 129–131] can be used to reconstruct an average cohort-level biomarker trajectory, which is continuous in contrast to the discrete description of the EBM. The models use short-term follow up biomarker measurements to provide samples of the gradient of a single common biomarker trajectory and integrate a differential equation to determine a best-fit or 'average' trajectory for the cohort. For example, Jack et al. [82] determine the time taken for amyloid accumulation to go from a normal to an abnormal level by

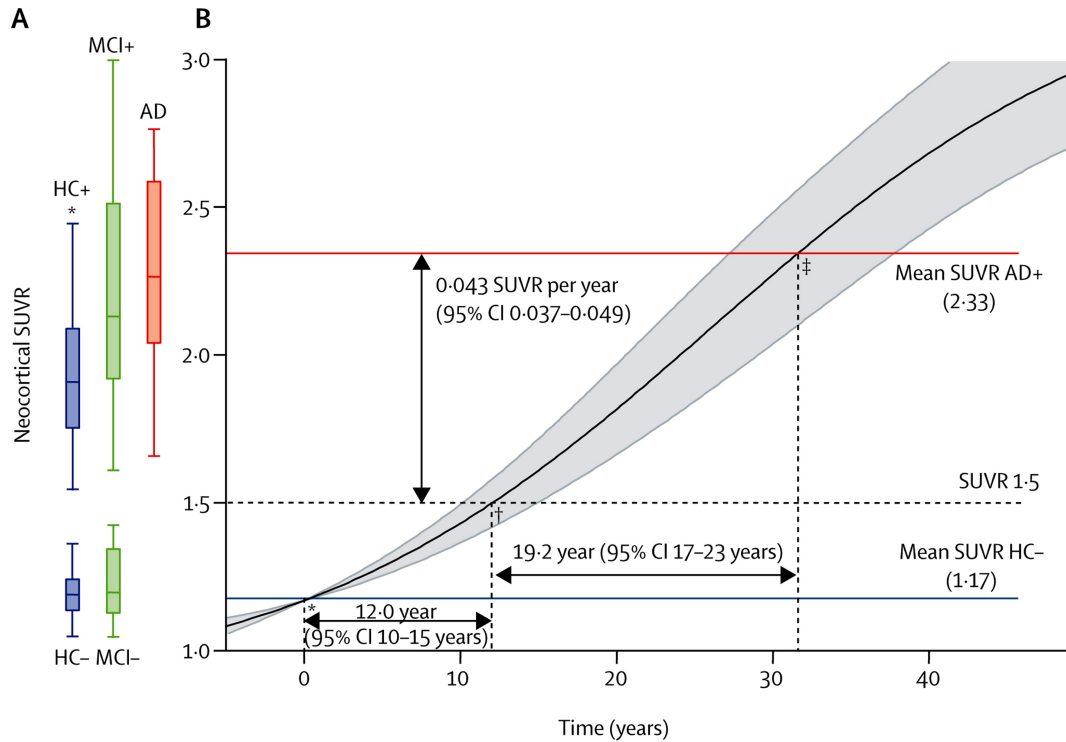


**Figure 2.2:** Sequence of events,  $\bar{S}$ , in dominantly-inherited Alzheimer's disease estimated by the event-based model proposed by Fonteijn et al. in 2012 [98]. Reprinted from [98], Copyright 2012, with permission from Elsevier.

fitting a DEM to data from serial amyloid-PET scans, finding that it takes approximately 15 years to go from a normal standard uptake value ratio (SUVR) of 1.5 to an abnormal SUVR of 2.5. Villemagne et al. [81] (Figure 2.3) perform a similar analysis to determine the time taken for several biomarkers to go from normal to abnormal, including amyloid-PET, hippocampal atrophy, episodic memory, gray matter volume and non-memory cognitive domains. DEMs have potential as a disease staging, monitoring and prognostic tool as they provide the rate of biomarker decline over the disease time course. Stochastic DEMs [131] can further express deviations from this average, providing prognostic information at the individual level. However, they model each biomarker individually, and so there is no guarantee of correspondence across disease stage and prognosis estimates between different biomarkers.

### 2.7.3 Self-modelling regression

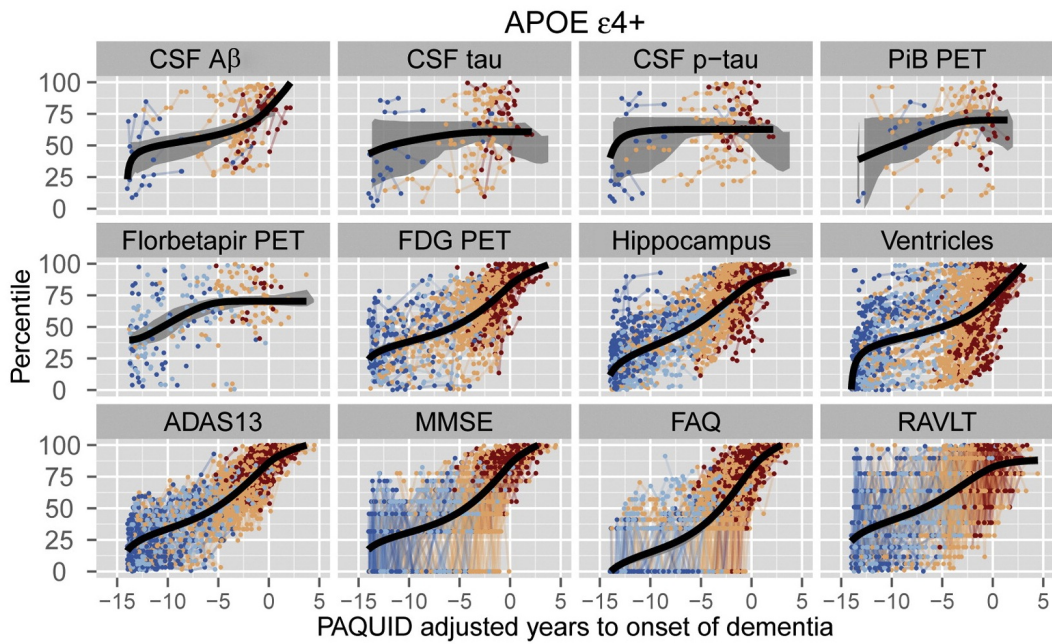
Self-modelling regression approaches [128, 132] bring together data from multiple biomarkers to estimate biomarker trajectories over a common disease timescale. Short-term follow up data from each individual provides samples of a common set of biomarker curves, which are used to estimate the population-level shape and rate



**Figure 2.3:** Differential equation model of amyloid deposition proposed by Villemagne et al. in 2013 [81]. The figure shows the estimated accumulation of amyloid plaques (measured using amyloid-PET imaging) with disease progression. Reprinted from [81], Copyright 2013, with permission from Elsevier.

of biomarker decline, as well as each individual's position and rate of decline. As with DEMs, the biomarker curves represent the average biomarker dynamics for a population. Donohue et al. [128] (Figure 2.4) use self-modelling regression to determine the trajectories of cognitive test scores, regional brain volumes from MRI, PET imaging measures, and CSF levels of amyloid-beta and tau. Jedynak et al. [132] formulate a similar model that uses cognitive test scores, CSF amyloid-beta and tau, and hippocampal volume on MRI to estimate a 'disease progression score', which is a continuous measure of disease stage that can be used as a time proxy. Bilgel et al. [133] have recently extended the 'disease progression score' model to work with voxelwise imaging data. However, this extension requires that biomarker trajectories are modelled as a linear function of disease stage, limiting the applicability of the model to multi-modality data where biomarkers may be dynamic at different points along the disease time course. Self-modelling regression approaches provide continuous disease staging, monitoring and prognostic measures that incorporate

information from multiple biomarkers. A key advantage of these models is that they provide a very complete picture of the disease, which can aid detailed disease understanding. Potential disadvantages are that they have many more parameters to estimate than simpler models like the EBM, so may be less stable; and the complex picture has a less straightforward interpretation than the discrete description, which may limit clinical utility.



**Figure 2.4:** Alzheimer's disease biomarker trajectories estimated using the self-modelling regression technique proposed by Donohue et al. in 2014 [128]. Reprinted from [128], Copyright 2014, with permission from Elsevier.

## 2.7.4 Critical assessment

To date, these data-driven models have shown compelling results that provide valuable insights into neurodegenerative disease progression patterns, particularly in AD. However, they remain an emerging area of research, and all the current models share a number of limitations and assumptions that are important to consider when interpreting results. One strong assumption that all the aforementioned models make is that all subjects follow a common progression pattern. Although some models allow for subjects to deviate from this common progression pattern, these deviations are assumed to be small, and none allow for subgroups of subjects that



follow completely different progression patterns. Such outliers are likely given the inherent heterogeneity of sporadic disease data sets, which contain some proportion of subjects with alternative neurodegenerative diseases, as well as mixed pathologies and a wide range of subject demographics. For this reason, practical applications of data-driven models often focus on more homogeneous population subgroups [81, 82, 128], for example subjects with increased genetic risk of developing the neurodegenerative disease of interest. Another assumption is the independence of biomarkers: although the models express temporal correlation of biomarker trajectories over the disease time course, they typically assume independence at any given time point. In practice, biomarkers often co-vary, for example amyloid-PET and CSF measures of amyloid-beta are measures of the same underlying pathology and are therefore strongly correlated. Failure to model this covariance tends to cause underestimation of the variance of progression patterns across the population. Data-driven models further assume that data is available from the full disease time course when in reality the data points may be sparse at the beginning and end of the disease progression, which may influence the estimation of biomarker trajectories.

Data-driven models are an emerging area of technology with major potential benefits to neurodegenerative disease research and clinical practice, and with wide potential further application to a range of other diseases or developmental processes. They can provide quantitative multi-modal pictures of the full disease time course for improved understanding of disease mechanisms to inform drug discovery; they naturally combine different types of information for earlier and more accurate differential diagnosis, and subject-specific prognostic information; they provide fine-grained staging scores or systems for more precise patient stratification supporting clinical trials for developing treatments and ultimately treatment deployment.

## 2.8 Summary

A wide range of models have been applied to AD. Neuropathological models and animal models have provided important insights into underlying disease mechanisms, but are not directly applicable for patient staging. Biomarker models have

been developed to allow subjects to be monitored in vivo. Simple scalar biomarker models are the most directly clinically applicable as they use well established biomarkers. However, their temporal resolution is limited by their reliance on a priori knowledge of a subject's disease stage, typically to just a few stages. High dimensional biomarker models have the potential to uncover more complex relationships within or between different data types, but again their temporal resolution is limited by their reliance on a priori disease staging. Data-driven models have been developed to allow biomarker trajectories to be recovered without a priori disease staging information. Such models have the potential to recover complex disease progression patterns and can be used as a fine-grained patient staging mechanism. However, they are still an emerging technology, and will require further validation and refinement before they are translated into a useful clinical tool.

## Chapter 3

# A data-driven model of biomarker changes in sporadic Alzheimer's disease

This chapter details sub-contributions (a) and (b) of contribution 1.3.1 (see Chapter 1, page 39). The work presented in this chapter was published in *Brain* and presented at the *Alzheimer's Association International Conference* in 2014. I developed the methodology, performed the analysis and wrote the manuscript; my co-authors provided feedback on the methodology and the manuscript.

### 3.1 Associated publications

A. L. Young, N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain*, 137(9):2564–2577, 2014

### 3.2 Introduction

Existing biomarkers of AD provide complementary information for disease staging and differential diagnosis. Determining the particular sequence and evolution of biomarker abnormality potentially provides a mechanism to stage and stratify patients throughout the full disease time course, and in particular, during the presymptomatic phase. This helps reduce heterogeneity in trial groups, match individuals

to putative treatments, and monitor treatment outcomes. Whilst new diagnostic criteria now incorporate biomarkers to allow earlier diagnosis [94], the evidence base for this is relatively limited. A major challenge of current AD research [75] is to construct models of disease progression that estimate biomarker ordering and dynamics directly from real-world datasets enabling quantitative evaluation of patient state.

The recently introduced event-based model (EBM) [98] provides a generative model of disease progression that can learn the ordering of biomarker changes from large cross-sectional (or short-term longitudinal to enable measurement of rates of atrophy) datasets, as well as providing insights into the uncertainty of the reconstructed ordering. The EBM defines the disease progression as a sequence of events at which individual biomarkers become abnormal. The EBM is probabilistic in the sense that it learns normal and abnormal distributions of biomarker values from the data, and so does not require a-priori staging or cut points. The EBM further enables the assignment of each subject to a disease stage. Previous work [98] demonstrated the EBM's ability to order biomarkers and generate staging measures derived from imaging data, in genetically defined disease and control populations (familial AD and Huntington's disease). However, the original EBM is not directly applicable to sporadic disease datasets, which have significant proportions of misdiagnosed cases in the patient group; and, particularly in AD research, a poorly defined control group because a significant number (estimated to be a third by the eighth decade) of apparently healthy elderly individuals have biomarker evidence consistent with presymptomatic AD [135, 136].

Here I reformulate the EBM for multi-modal data from a heterogeneous sporadic disease population. The new EBM accommodates a modest proportion of misdiagnosed patients as well as allowing for presymptomatic cases contaminating the control group. I apply this EBM to the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to obtain characteristic biomarker orderings from various subgroups, as well as their uncertainty. I demonstrate the fine-grained staging potential of the EBM and its ability both to classify cognitively normal (CN) and AD

subjects and to predict conversion from mild cognitive impairment (MCI) to AD and CN to MCI.

## 3.3 Materials and methods

### 3.3.1 Data description

#### 3.3.1.1 Subjects

I downloaded data from LONI<sup>1</sup> on 5 February 2013, and included all 285 subjects (CN, MCI or AD) that had a CSF examination at baseline, standardised cognitive assessment at baseline<sup>2</sup>, which included: the Mini Mental State Examination (MMSE) [85], the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) [137] (modified 13 item ADAS-Cog, which omits item 13), and the Rey Auditory Verbal Learning Test (RAVLT) [138] (immediate recall score, i.e. the sum of trials 1 to 5), and useable 1.5T MRI imaging at baseline and 1 year. Clinical diagnosis (CN/MCI/AD) was also recorded. Other possible biomarkers, e.g. FDG PET and amyloid PET, were not included in the present analysis because they limit the number of available subjects: less than half of the subjects with CSF and MRI data at baseline underwent an FDG PET scan at baseline, and very few had baseline amyloid PET imaging. CSF measures of  $A\beta_{1-42}$ , p-tau and t-tau were performed centrally, as previously described [139]. The CSF t-tau and p-tau data were log transformed to improve normality. I downloaded APOE genotype, for which methods have been published previously [140], for each individual from the LONI website. For validation of the staging system derived from the EBM, I downloaded the aforementioned set of imaging, clinical and CSF data at 12 and 24 month follow up time points. CSF values are known to only be comparable when they are all processed at the same time. Therefore I downloaded longitudinal CSF data collected over 4 years that had been reprocessed at the end of the 4 years, so as to obtain baseline, 12 and 24 month CSF data which were processed in the same batch. As an outcome measure, I downloaded clinical diagnoses at all available time points up

---

<sup>1</sup>[www.loni.ucla.edu/ADNI/](http://www.loni.ucla.edu/ADNI/)

<sup>2</sup>For details see [www.adni-info.org/Scientists/Pdfs/adniproceduresmanual12.pdf](http://www.adni-info.org/Scientists/Pdfs/adniproceduresmanual12.pdf).

to 72 months.

### 3.3.1.2 MRI methods

Details of the MRI methodology have previously been described [141]. Cross-sectional regional measures of brain volumes known to be affected in AD: the hippocampus, entorhinal cortex, middle temporal gyrus, fusiform, ventricles and whole brain, as well as total intracranial volume (TIV), were calculated at baseline using FreeSurfer Version 4.3, which is documented and freely available for download online<sup>3</sup>. All regional volumes were normalised by dividing by TIV for each subject.

Longitudinal measures of regional volume change between 0 and 12 months were obtained using the boundary shift integral (BSI): volume change was measured for the whole brain using the KN-BSI method [142], and for the hippocampus using the MAPS-HBSI method [143].

### 3.3.1.3 Event set

The biomarkers available for all the subjects provide the following set of 14 biomarker transition 'events', each of which corresponds to a biomarker becoming abnormal, i.e. changing from the 'control' to 'AD' state

- Three CSF events:  $A\beta_{1-42}$  (Abeta), P-tau and T-tau
- Three cognitive events: ADAS-Cog, RAVLT and MMSE
- Six regional brain volume events: Brain, Ventricles, Hippocampus, Entorhinal, Mid Temporal and Fusiform volumes
- Two rates of atrophy events: rates of Hippocampal and Brain Atrophy

## 3.3.2 Event sequences

I defined four population subgroups:

- Whole population: all subjects.
- Amyloid positive ( $A\beta+$ ): subjects with CSF  $A\beta_{1-42} < 192$  pg/ml. This cut point was chosen according to the results of Shaw et al. [139] who determined

---

<sup>3</sup><http://surfer.nmr.mgh.harvard.edu/>

cut points using a maximum accuracy classification of autopsy confirmed AD and CN subjects.

- APOE4 positive (APOE4+): subjects with one or more APOE4- $\epsilon$ 4 alleles.
- Amyloid positive APOE4 positive ( $A\beta$ +APOE4+): subjects who are both  $A\beta$ + and APOE4+.

### 3.3.2.1 The event-based model

I estimated the most likely ordering of events and its uncertainty in each subgroup using the EBM [98]. The EBM treats each biomarker as either ‘normal’, i.e. non-pathological, or ‘abnormal’, i.e. as seen in AD. The switch from normal to abnormal is termed an ‘event’. The occurrence of any particular event,  $E_i$ ,  $i = 1 \dots I$ , is informed by the corresponding measurements  $x_{ij}$  of biomarker  $i$  in subject  $j$ ,  $j = 1 \dots J$ . The whole dataset  $X = \{x_{ij} | i = 1 \dots I, j = 1 \dots J\}$  contains measurements of each biomarker in each subject. The most likely ordering of the events is the sequence  $S$  that maximises the data likelihood

$$P(X|S) = \prod_{j=1}^J \left[ \sum_{k=0}^I \left( P(k) \prod_{i=1}^k P(x_{ij}|E_i) \prod_{i=k+1}^I P(x_{ij}|\neg E_i) \right) \right], \quad (3.1)$$

where  $P(x|E_i)$  and  $P(x|\neg E_i)$  are the likelihoods of measurement  $x$  given that biomarker  $i$  has or has not become abnormal, respectively, and  $P(k)$  is the prior likelihood of being at stage  $k$ , i.e. events  $E_1, \dots, E_k$  have occurred, and events  $E_{k+1}, \dots, E_I$  have yet to occur, which I assume is uniform. This uniform prior assumes no knowledge of any patient’s disease stage a-priori, which imposes the least information possible on estimated orderings.

In addition to finding the most likely sequence,  $P(X|S)$  can be evaluated for any sequence to establish the relative likelihood of all sequences. This provides insight into the uncertainty of the ordering. The positional variance diagram [98] (Figure 3.1 A-D) visualises both the maximum likelihood sequence and its uncertainty by plotting the likelihood that each event appears in each position in the sequence, i.e. the entry of each position is  $\sum_{S \in S_{ik}} P(X|S)$  where  $S_{ik}$  is the set of all sequences with

event  $i$  at position  $k$ .

### 3.3.2.2 Model of the event distribution

Evaluation of Equation (3.1) requires models for each of the event distributions,  $P(x|E_i)$  and  $P(x|\neg E_i)$ . The original EBM in [98] used a familial AD dataset for which the control group was well defined allowing direct estimation of  $P(x|\neg E_i)$ . In sporadic AD, however, a significant proportion of the CN control group may have presymptomatic AD. To counter this, I reformulated the EBM approach so that the event distributions are estimated by fitting a mixture of two normal distributions (a gaussian mixture model). I fitted this mixture of normal distributions to each biomarker separately using data from all subjects to obtain the parameters of the two models. To ensure a robust fit, particularly for biomarkers where the distributions of the healthy and diseased population overlap significantly, I constrain the standard deviations so that the standard deviation of  $P(x|\neg E_i)$  and  $P(x|E_i)$  is less than or equal to that of the CN and AD group respectively. This is a weak constraint designed simply to guide the mixture model away from physically unrealistic solutions. Importantly, whilst this modelling approach can be used to determine fixed cut points for each biomarker, the model here is not dependent on these cut-points, using a probability function to determine the most likely sequencing of event switches.

### 3.3.2.3 Summary of estimation procedure for the EBM

The model fitting procedure for the EBM is as follows.

1. Fit a mixture model to the data for all subjects to estimate the parameters of the event distributions,  $P(x|E_i)$  and  $P(x|\neg E_i)$ .

For each population subgroup:

2. Find the characteristic event sequence  $\bar{S}$  that maximises the data likelihood  $P(X|S)$  by performing a greedy ascent algorithm.
3. Take Markov Chain Monte Carlo (MCMC) samples of the data likelihood  $P(X|S)$ , initialised from the maximum likelihood event sequence  $\bar{S}$ , to estimate the uncertainty in the characteristic event sequence.



### 3.3.2.4 Implementation of EBM estimation procedure

All experiments were performed using Matlab on a standard workstation (Intel Core i7, 3.1 GHz, 8GB memory). The event distribution (gaussian mixture model) parameters were optimised using the ‘fmincon’ constrained optimisation solver with the ‘sqp’ - sequential quadratic programming - method. All of the gaussian mixture model fits were assessed visually to ensure the solutions were physically plausible. To find the characteristic event sequence  $\bar{S}$ , I performed 10,000 iterations of a greedy ascent algorithm (as in [98]), initialised from 25 random start points. I checked that the start points converged to a single maximum to ensure the global optimum had been found. I then ran an MCMC algorithm (as in [98]) to draw samples from the posterior distribution  $P(S|X)$ . I ran the MCMC algorithm for 100,000 iterations, checking that the MCMC trace showed good mixing properties. I initialised the MCMC algorithm at  $\bar{S}$  so a burn-in period was not required.

The computational complexity of the gaussian mixture model fitting is  $O(I)$ , where  $I$  is the number of biomarker events, and the full optimisation procedure for each biomarker takes less than 0.2 seconds. The computational time for the greedy ascent and MCMC algorithms depends on (i) the time taken to compute the data likelihood  $P(X|S)$ , and (ii) the number of samples required for the algorithm, i.e. for the greedy ascent algorithm to reach the maximum likelihood solution, or for the MCMC chain to sample  $P(S|X)$ . Calculating the value of  $P(X|S)$  involves computing a product over all  $I$  biomarker events, summing this over all  $I + 1$  possible stages in the sequence, and then taking the product over all  $J$  subjects, which is of order  $O(I^2J)$ , as  $P(x_{ij}|\neg E_i)$  and  $P(x_{ij}|E_i)$  can be computed ahead of time. Although the space of all possible sequences has a size of  $I!$ , the maximum pairwise distance between any pair of sequences is  $\frac{I(I-1)}{2}$ . This means that the number of iterations required for convergence of the greedy ascent algorithm, and for the MCMC chain to sample  $P(S|X)$ , should scale with approximately  $O(I^2)$ . For the set of biomarker events used here, running 10,000 iterations of the greedy ascent algorithm takes approximately 1.5 seconds, and taking 100,000 samples for the MCMC algorithm takes approximately 15 seconds (i.e. computing  $P(X|S)$  takes approximately 1.5

$\times 10^{-4}$  seconds).

### 3.3.2.5 Cross-validation of the event sequence

I performed cross-validation of the maximum likelihood event sequence returned by the EBM (Figure 3.1 E-H) by re-estimating the event distributions and maximum likelihood sequence (Figure 3.1 A-D) for 100 bootstrap samples of the data. The positional variance diagrams for the cross validation results show the proportion of bootstrap samples in which event  $i$  appears at position  $k$  of the maximum likelihood sequence.

## 3.3.3 Patient staging

Once the characteristic sequence  $\bar{S}$  has been determined using the EBM, the simplest way to assign a stage for a particular subject, which I adopt here, is to find the stage which is assigned the highest probability by the model, i.e. the stage,

$$\operatorname{argmax}_k P(X_j, k | \bar{S}) = \operatorname{argmax}_k P(k) \prod_{i=1}^k P(x_{ij} | E_i) \prod_{i=k+1}^I P(x_{ij} | \neg E_i), \quad (3.2)$$

that maximises the probability of the data given the maximum likelihood event sequence. As before, I make no a-priori assumptions about model stage by assuming the prior,  $P(k)$ , is uniform. The stage ranges from 0 to  $I$  (the number of events). Thus the idealised model for stage  $k$  is that all events up to and including  $k$  have occurred and the events after  $k$  have not occurred. However, the assignment of stage  $k$  to a particular patient does not mean they fit the model exactly; it is simply the stage most compatible with their measurements.

### 3.3.3.1 Longitudinal validation

To assess the consistency of patient staging measures longitudinally (Figure 3.3) I evaluated each patient's stage at all follow up time points which met the inclusion criteria: subjects had to have measurements for all biomarkers, including an MRI scan 12 months later in order to calculate the BSI over a consistent time frame. There were two follow up time points which met these criteria: 12 months (Figure 3.3 A) and 24 months (Figure 3.3 B). I compared each subject's EBM stage at follow

up with their baseline EBM stage, which was re-evaluated using the reprocessed CSF measures so as to ensure that the CSF was processed consistently for all time points.

### 3.3.3.2 Prediction of conversion

Patient staging derived from the EBM can be used to predict conversion from MCI to AD or CN to MCI (Table 3.2) by categorising subjects according to their EBM stage at baseline. I performed a binary classification of subjects into MCI-stable and MCI-converters, and CN-stable and CN-converters, by thresholding on patient EBM stage. Stable subjects were defined as those with an MCI or CN diagnosis who remained with an MCI or CN diagnosis at the end of a 12, 24, 36, 48 or 60 month follow up period. Converters were defined as those with an MCI or CN diagnosis who were diagnosed with AD or MCI, respectively, at the end of a 12, 24, 36, 48 or 60 month follow up period. I used the EBM stage that maximises balanced accuracy to classify subjects. Balanced accuracy is the average of the sensitivity and specificity, which is similar to accuracy but does not depend on disease prevalence. To test the effect of increasing EBM stage on the probability of conversion from MCI to AD and CN to MCI (Table 3.3 and Figure 3.4) I used Cox proportional hazards Models where the event was conversion to AD or MCI respectively and the input variables were patient EBM stage and demographic factors: age, sex, education and APOE4 carrier status (presence of an APOE4 allele). Time to event data for subjects who did not convert was considered censored at their last available diagnosis. Statistical significance was set at  $P < 0.05$ .

### 3.3.4 Staging using cross-sectional data alone

To demonstrate the EBM's ability to stage patients using purely cross-sectional measures I repeated the patient staging by fitting the EBM for a subset of 12 events (Tables 3.4, 3.5, 3.6, Figures 3.5, 3.6, 3.7, 3.8), excluding atrophy rates. The inclusion criteria were the same as used previously except follow up MRI scans at 12 months were not required. As before, patient staging results were evaluated for the whole population using the maximum likelihood event sequence determined over

all subjects, but with atrophy rates removed (Figure 3.5 A).

## 3.4 Results

### 3.4.1 Subjects

Study subject demographics are summarised in Table 3.1. Of the 285 subjects that met the inclusion criteria, 189 were amyloid positive ( $A\beta+$ ), 139 were APOE4 positive (APOE4+), and 123 were amyloid positive and APOE4 positive ( $A\beta+$ +APOE4+).

	Demographics	CN	MCI	AD
All subjects	N	92	129	64
	Sex M/F	48/44 (52%)	82/47 (64%)	34/30 (53%)
	Age	75 $\pm$ 5	73 $\pm$ 7	75 $\pm$ 8
	Education	15.6 $\pm$ 2.9	15.9 $\pm$ 3	15 $\pm$ 3
	APOE4 +/-	22/70 (24%)	72/57 (56%)	45/19 (70%)
$A\beta+$	N	34	96	59
	Sex M/F	19/15 (56%)	58/38 (60%)	31/28 (53%)
	Age	76 $\pm$ 5	73 $\pm$ 7	74 $\pm$ 8
	Education	15.8 $\pm$ 3.3	15.7 $\pm$ 3.1	15 $\pm$ 3.1
	APOE4 +/-	15/19 (44%)	63/33 (66%)	45/14 (76%)
APOE4+	N	22	72	45
	Sex M/F	15/7 (68%)	39/33 (54%)	25/20 (56%)
	Age	75 $\pm$ 6	73 $\pm$ 6	75 $\pm$ 7
	Education	15.6 $\pm$ 3.4	15.8 $\pm$ 2.9	14.6 $\pm$ 3
	APOE4 +/-	22/0 (100%)	72/0 (100%)	45/0 (100%)
$A\beta+$ +APOE4+	N	15	63	45
	Sex M/F	10/5 (67%)	35/28 (56%)	25/20 (56%)
	Age	77 $\pm$ 6	73 $\pm$ 6	75 $\pm$ 7
	Education	15.5 $\pm$ 3.8	15.8 $\pm$ 2.9	14.6 $\pm$ 3
	APOE4 +/-	15/0 (100%)	63/0 (100%)	45/0 (100%)

**Table 3.1:** Baseline demographics for the whole population and population subgroups. Age and education are in years (mean  $\pm$  standard deviation). Reproduced from [134]. CC BY 3.0.

### 3.4.2 Event sequences

Figures 3.1 A-D show positional variance diagrams for each population subgroup. Each positional variance diagram shows the maximum likelihood event sequence and its uncertainty. Figures 3.1 E-H show positional variance diagrams obtained

from cross-validation of the maximum likelihood ordering.

The event sequences in all four populations (Figure 3.1 A-D) showed broad agreement with hypothetical models such as Jack et al. [74]: CSF biomarkers were shown to be early events, followed by atrophy rates, then cognitive test scores and hippocampal and entorhinal volume, and finally other regional brain volumes. Cross-validation (Figure 3.1 E-H) confirmed high confidence in the ordering of these sets of events: for all populations, the ordering strongly placed CSF and atrophy rates before cognitive test scores and hippocampal and entorhinal volume, and the remaining regional volume changes last.

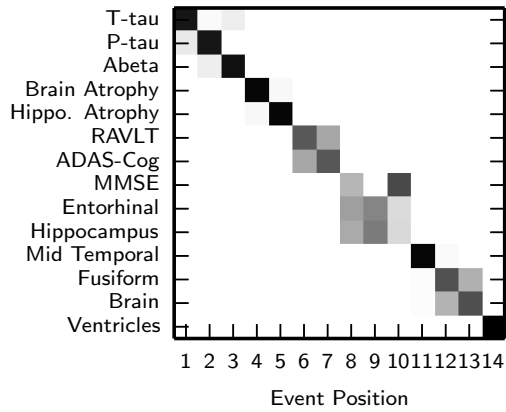
#### 3.4.2.1 Whole population

The maximum likelihood ordering for the whole population (Figure 3.1 A) showed some departures from current thinking in neurology [74], although the uncertainty was high (Figure 3.1 E). First, CSF t-tau occurred prior to p-tau. It might be expected that p-tau is an earlier marker of AD than t-tau [75], being a more specific measure of the build up of NFTs than t-tau [62], which measures associated neuronal damage. Second, both t-tau and p-tau occurred before  $A\beta_{1-42}$ , whereas amyloid plaque deposition is widely considered to be the initiating event in AD [144]. Third, brain atrophy rate came before hippocampal atrophy rate, which is at odds with the findings of MRI regional atrophy rate studies (e.g. [114]).

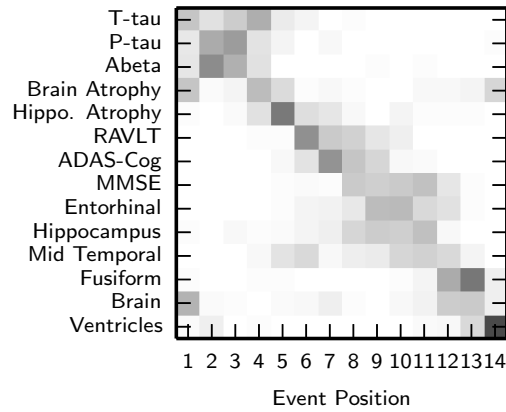
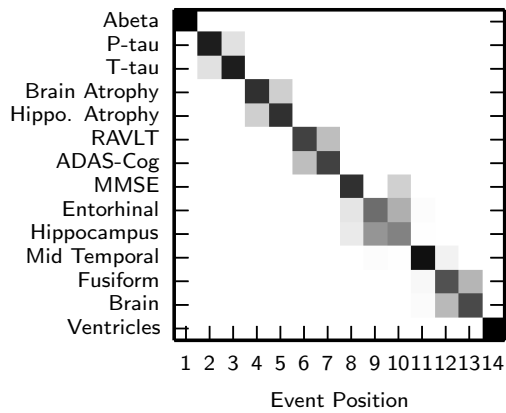
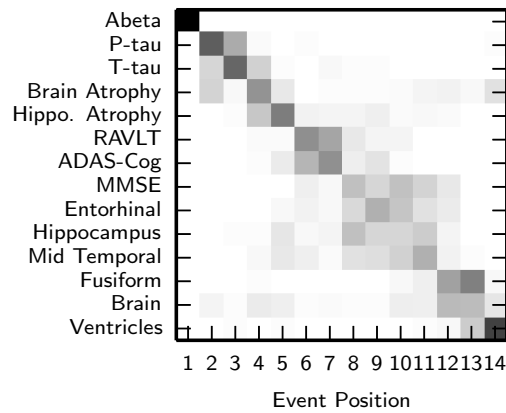
#### 3.4.2.2 $A\beta+$ and APOE4+ subjects

The  $A\beta+$ , APOE4+ and  $A\beta$ +APOE4+ groups (Figures 3.1 B-D) showed a distinct ordering of the CSF biomarkers:  $A\beta_{1-42}$ , p-tau, t-tau, which replicated the ordering described by hypothetical models [74,75]. Cross-validation (Figure 3.1 F-H) of the event sequence in these groups showed a much greater confidence in the ordering of CSF biomarkers compared to the whole population (Figure 3.1 E), which is more heterogeneous. In the  $A\beta+$  group (Figure 3.1 B), brain atrophy rate was ordered before hippocampal atrophy rate, but the ordering was weaker than the whole population. In the APOE4+ and  $A\beta$ +APOE4+ groups (Figures 3.1 C-D) hippocampal atrophy rate clearly occurred before brain atrophy rate.

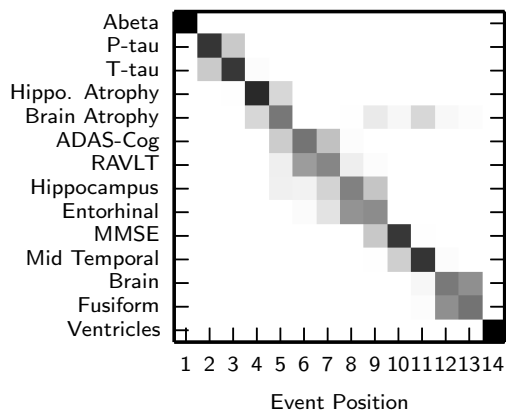
A. Whole population



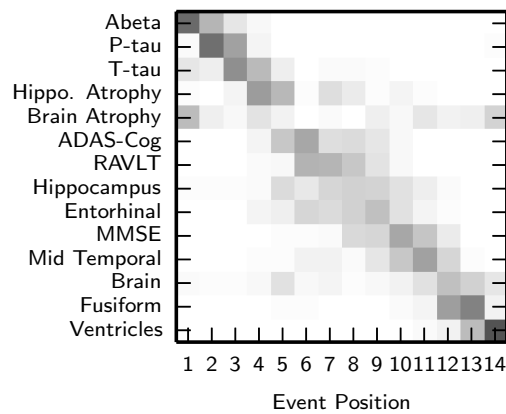
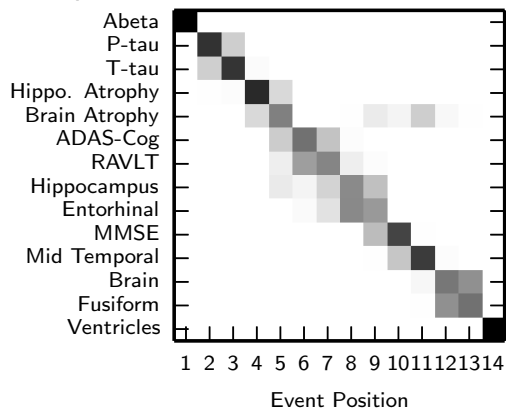
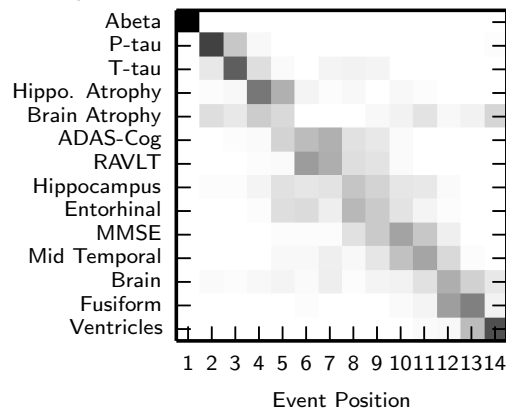
E. Whole population

B.  $A\beta+$ F.  $A\beta+$ 

C. APOE+



G. APOE+

D.  $A\beta+$  APOE+H.  $A\beta+$  APOE+

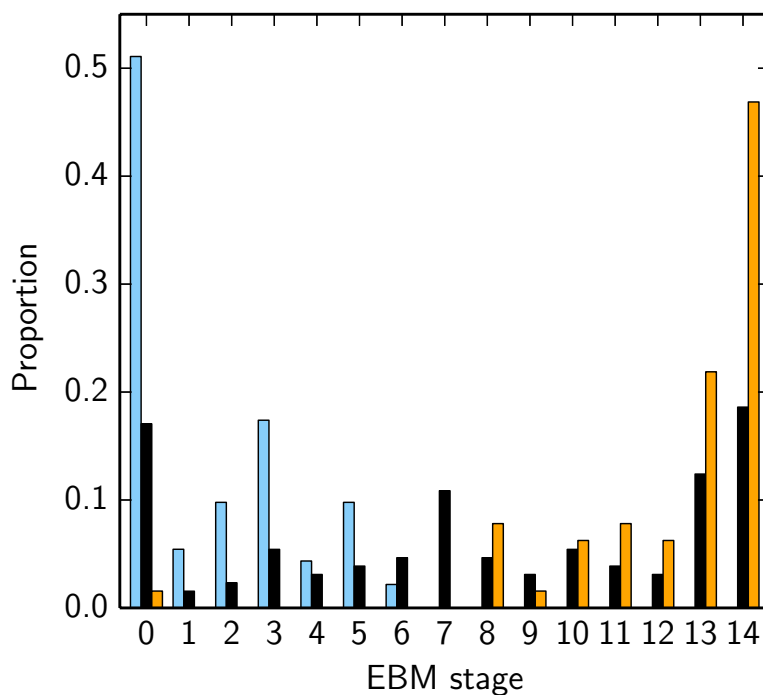
**Figure 3.1:** Positional variance diagrams showing the distribution of event sequences in population subgroups. (A-D) Positional variance diagrams of the uncertainty in the maximum likelihood event ordering estimated by taking MCMC (Markov chain Monte Carlo) samples using the EBM. (E-H) Positional variance diagrams from cross-validation of the maximum likelihood event sequence by bootstrap resampling of the data. These diagrams overestimate the uncertainty, giving a more conservative picture than the left hand column. Each entry in the positional variance diagram represents the proportion of MCMC samples, in A-D, or bootstrap samples, in E-H, in which events appear at a particular position in the sequence (x-axis). This proportion ranges from 0 in white to 1 in black. The y-axis orders events by the maximum likelihood sequence. Where rows have a single black block on the diagonal, such as the top five events in the diagram for the whole population, the ordering is strong and permutations of those events are unlikely. Grey blocks, such as the Mini-Mental State Examination (MMSE) score, entorhinal volume and hippocampal volume in the whole population, show that permuting the order of the events has little effect on the likelihood so their ordering is weak.  $A\beta+$  = amyloid+; Abeta = amyloid- $\beta$ ; P-tau = phosphorylated tau; T-tau = total tau; RAVLT = Rey Auditory Verbal Learning Test. Reproduced from [134]. CC BY 3.0.

### 3.4.3 Patient staging

#### 3.4.3.1 Cross-sectional distribution of stages

Figure 3.2 shows the distribution of patient stages for the whole population. All patient staging results were evaluated for the whole population using the maximum likelihood event sequence determined over all subjects (Figure 3.1 A). The distributions of EBM stages for CN and AD subjects were strongly separated and thresholds at middle stages classify CN vs. AD with accuracy greater than 99%. The majority of CN subjects have no biomarker abnormalities, so are assigned stage 0, or abnormalities only in CSF, so are assigned stages 1-3. A small number of CN subjects also showed rates of atrophy events, so are assigned stages 4-6. Most AD subjects have abnormal CSF, atrophy rate, cognitive symptoms and low hippocampal and entorhinal volume so are assigned later stages. The majority of AD subjects were assigned the final stage in the progression, showing that the model configuration that fits their data best is where all of the events have occurred. The distribution of MCI stages overlapped with the distribution of stages for CN and AD subjects but with a greater concentration of subjects around the middle stages, suggesting that

these subjects show CSF abnormalities, abnormal rates of atrophy, and some cognitive symptoms. To explore the extent to which choice of cognitive test affects the staging (and event sequence) output, I assessed the effect of adding in an additional memory test, the Logical Memory II subscale (delayed paragraph recall) from the Wechsler Memory Scale - Revised. Results (not shown) confirm that using this additional cognitive test score provides a very similar distribution of patient EBM stages, with logical memory occurring immediately prior to the RAVLT in the event sequence.



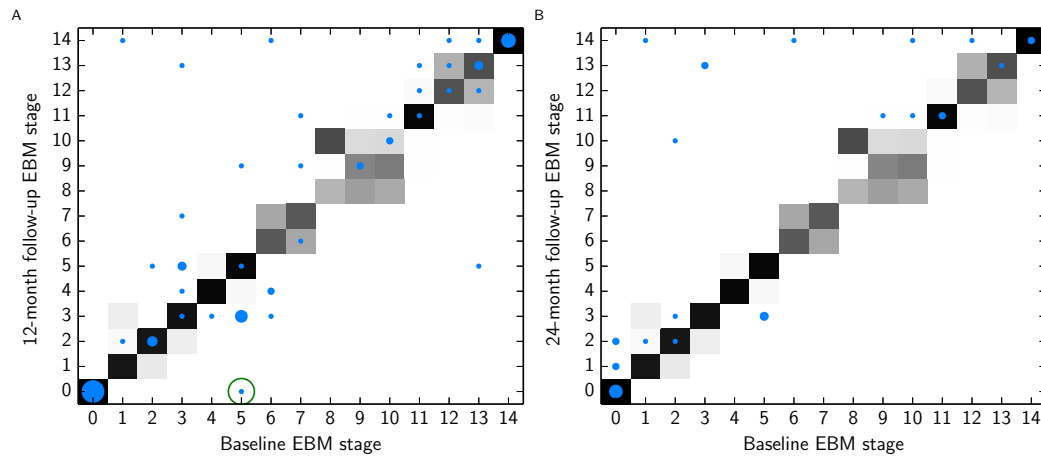
**Figure 3.2:** Proportion of patients in each diagnostic category at each EBM stage. Proportion of cognitively normal in light blue, mild cognitive impairment in black, and Alzheimer's disease in orange. Each EBM stage on the x-axis corresponds to the occurrence of a new biomarker transition event. Stage 0 corresponds to no events having occurred and stage 14 is when all events have occurred. Events are ordered by the maximum likelihood event sequence for the whole population as shown in Fig. 3.1 A. Reproduced from [134]. CC BY 3.0.

### 3.4.3.2 Longitudinal consistency

Figure 3.3 compares each subject's EBM stage at baseline with their EBM stage at 12 and 24 month follow ups. Patient staging showed good longitudinal consistency, with the EBM stage of each subject generally increasing or remaining stable at each



follow up (most points are within or above the grey shaded area, which represents the uncertainty estimated by the EBM, as shown in figure 3.1 A). The small number of individuals whose EBM stage decreased longitudinally (below the diagonal) by more than the uncertainty estimated by the EBM (shaded in grey) were all subjects who improved from an abnormal to a normal score on one or more of the three cognitive tests (MMSE, RAVLT, and ADAS-Cog) and/or two atrophy rates (brain atrophy rate and hippocampal atrophy rate) with the exception of one subject (circled in green) whose CSF  $A\beta_{1-42}$  levels increased from a clearly abnormal level of 139 pg/ml at baseline to a more borderline level of 207 pg/ml at the 12 month follow up.



**Figure 3.3:** Longitudinal consistency of patient staging in the whole population over a (A) 12-month and (B) 24-month follow-up period. The size of the dot plotted at each point corresponds to the number of subjects with that particular baseline and follow-up EBM stage. The largest dot, at (0,0) represents 19 subjects in A and seven subjects in B, and the smallest dots represent one subject. The grey shaded area visualizes the uncertainty in the sequence estimated by the EBM (as shown in Fig. 3.1 A). Subjects whose EBM stage is longitudinally consistent are on or above the line  $y = x$  and/or within the grey shaded area. Subjects whose CSF levels (CSF amyloid- $\beta_{1-42}$  and/or phosphorylated tau and/or total tau) change from an abnormal to a normal level at follow-up are circled in green. Reproduced from [134]. CC BY 3.0.

### 3.4.3.3 Prediction of clinical outcomes

Table 3.2 A shows the balanced accuracy, sensitivity, specificity, area under the ROC curve (AUC), and maximum accuracy threshold EBM stage for classification of MCI-stable vs. MCI-converters over different follow up durations. The balanced

accuracy and AUC of the classification were comparable to state of the art classification techniques [121]. As the duration of the follow up increased, the maximum balanced accuracy threshold decreased, i.e. later EBM stages were better at predicting faster conversion times. These optimal stage thresholds suggest that abnormal CSF measures, atrophy rate, cognitive test scores and hippocampal and entorhinal volume provide the best prediction of conversion in 2 years or less, whereas just abnormal CSF, atrophy rate and ADAS-Cog and RAVLT scores is the combination that best predicts conversion over a period of 3 to 5 years.

The same statistics are shown in Table 3.2 B for classification of CN-stable vs. CN-converters. Again the threshold EBM stage decreased for increasing follow up durations, with abnormal CSF t-tau, p-tau and  $A\beta_{1-42}$  levels best predicting conversion from CN to MCI over a period of 4 years or less, but just abnormal CSF t-tau and p-tau best predicting conversion over 5 years.

#### A. MCI-stable vs. MCI-converters

	Balanced Acc. (%)	Sensitivity (%)	Specificity (%)	AUC	Threshold Stage	N-c/N-s
12 months	67	60	73	0.69	12	30/96
24 months	68	57	80	0.71	12	53/64
36 months	77	86	69	0.78	7	65/48
48 months	78	83	72	0.76	7	70/18
60 months	76	84	69	0.77	7	73/16

#### B. CN-stable vs. CN-converters

	Balanced Acc. (%)	Sensitivity (%)	Specificity (%)	AUC	Threshold Stage	N-c/N-s
12 months	84	100	68	0.76	3	2/90
24 months	66	67	66	0.62	3	6/83
36 months	68	63	73	0.62	3	8/73
48 months	66	58	74	0.65	3	12/49
60 months	76	75	76	0.75	2	16/38

**Table 3.2:** Classification results for discriminating MCI-stable versus MCI-converters and CN-stable versus CN-converters using patient stage at baseline. CN = cognitively normal; MCI = mild cognitive impairment. Reproduced from [134]. CC BY 3.0.

Table 3.3 shows the hazard ratio and statistical significance of each variable in the Cox proportional hazards models. Increasing EBM stage was a significant

hazard for conversion from both MCI to AD and CN to MCI. Figure 3.4 shows the estimated probability of remaining CN or MCI depending on baseline EBM stage.

#### A. MCI to AD progression

	HR (CI)	p-value	Corrected HR (CI)	Corrected p-value
EBM Stage	1.15 (1.09-1.21)	$1.58 \times 10^{-7*}$	1.15 (1.09-1.21)	$2.06 \times 10^{-7*}$
Age	0.99 (0.96-1.03)	0.68	0.99 (0.96-1.02)	0.49
Education	0.98 (0.91-1.05)	0.55	0.98 (0.90-1.05)	0.51
APOE4 Carrier	1.55 (0.97-2.48)	0.065	1.19 (0.73-1.94)	0.49
Male	0.77 (0.49-1.23)	0.28	0.85 (0.50-1.45)	0.55

#### B. CN to MCI progression

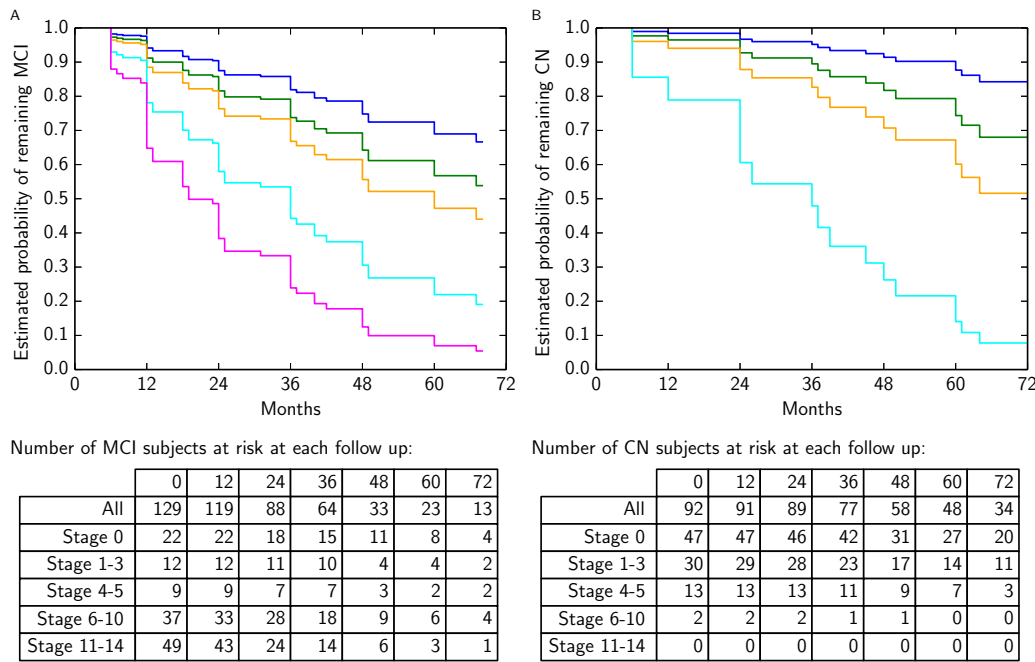
	HR (CI)	p-value	Corrected HR (CI)	Corrected p-value
EBM Stage	1.34 (1.07-1.69)	0.012*	1.31 (1.02-1.68)	0.033*
Age	0.99 (0.90-1.09)	0.84	0.98 (0.89-1.08)	0.67
Education	1.03 (0.88-1.22)	0.69	1.02 (0.86-1.20)	0.83
APOE4 Carrier	3.15 (1.19-8.30)	0.021*	2.47 (0.85-7.17)	0.096
Male	1.75 (0.65-4.74)	0.27	1.45 (0.49-4.28)	0.5

**Table 3.3:** Hazard ratios with 95% confidence intervals (CI) for conversion from MCI to AD, and CN to MCI, obtained by fitting uncorrected and corrected Cox proportional hazards models. \* $P < 0.05$ . CN = cognitively normal; MCI = mild cognitive impairment; AD = Alzheimer's disease. Reproduced from [134]. CC BY 3.0.

### 3.4.4 Staging using cross-sectional data alone

I repeated all analyses for purely cross-sectional measures, i.e. excluding rates of atrophy, to demonstrate the clinical application of the EBM's staging system, where patients need to be staged at one point in time. Table 3.4 gives demographic information for the 325 subjects that met the inclusion criteria, of which 216 were  $A\beta+$ , 159 were APOE4+, and 141 were  $A\beta+$ +APOE4+.

Removing atrophy rates had little effect on biomarker ordering (Figure 3.5) or the cross-sectional distribution (Figure 3.6) and longitudinal consistency (Figure 3.7) of staging. Again, individuals whose EBM stage decreased longitudinally (below the diagonal) by more than the uncertainty estimated by the EBM (shaded in grey) improved from a clearly abnormal to a more normal score on one or more of the three cognitive tests (MMSE, RAVLT, and ADAS-Cog) with the exception of



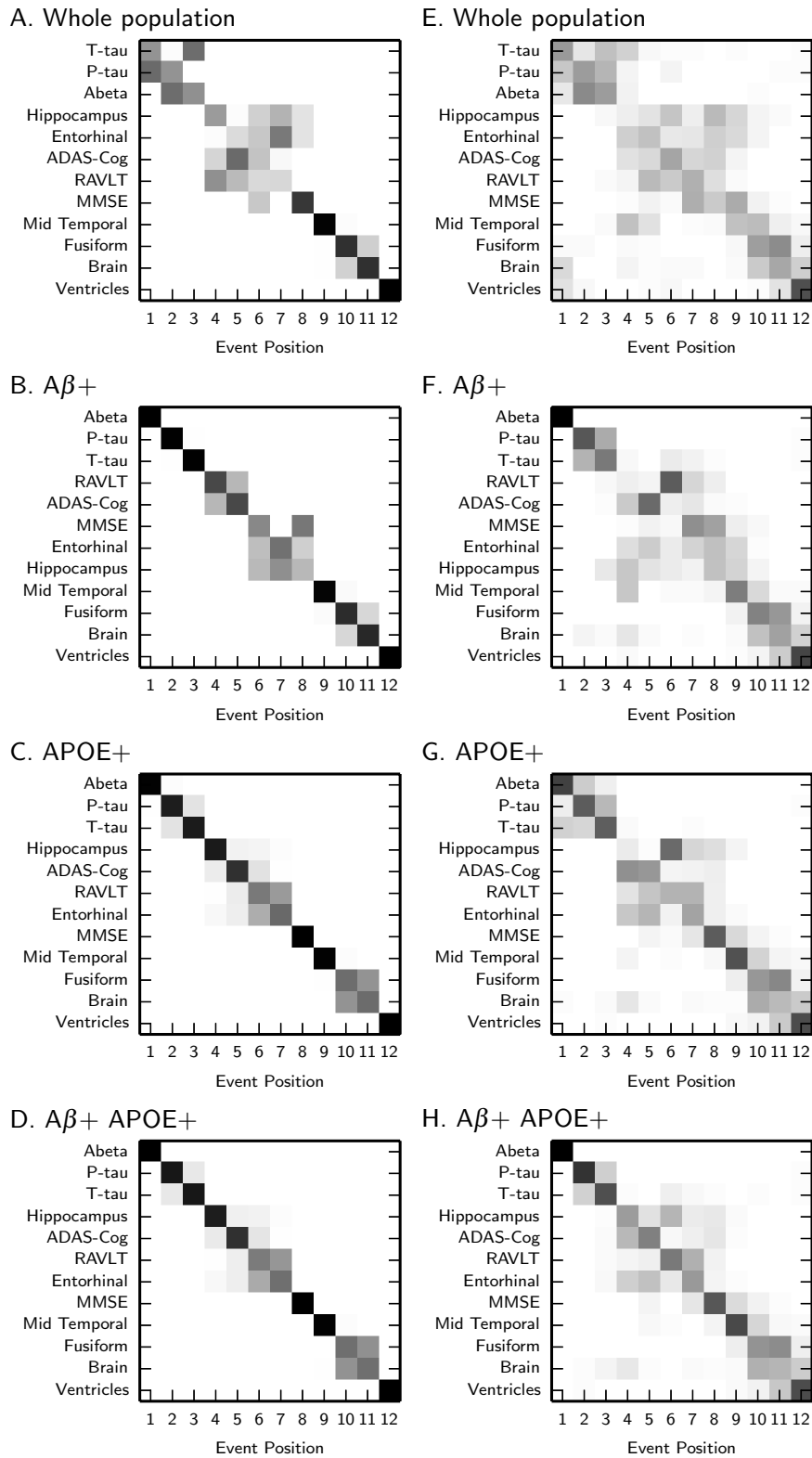
**Figure 3.4:** Estimated probability of remaining (A) MCI or (B) CN for different baseline EBM stages, obtained by fitting Cox proportional hazards models. These estimated probabilities are shown for the average population demographics (74.1 years of age, 15.6 years of education, APOE4 negative, male sex). Stages are grouped so that normal (blue) = stage 0, CSF (green) = stages 1-3, atrophy (orange) = stages 4-5, cognition (cyan) = stages 6-10, which includes hippocampal and entorhinal volume as well as cognitive test scores, volume (magenta) = stages 11-14. The table details the number of subjects at risk at each follow-up time point. MCI = mild cognitive impairment; CN = cognitively normal. Reproduced from [134]. CC BY 3.0.

two subjects (circled in green) whose CSF levels (CSF  $A\beta_{1-42}$  and/or p-tau and/or t-tau) changed from an abnormal to a more normal level at follow up.

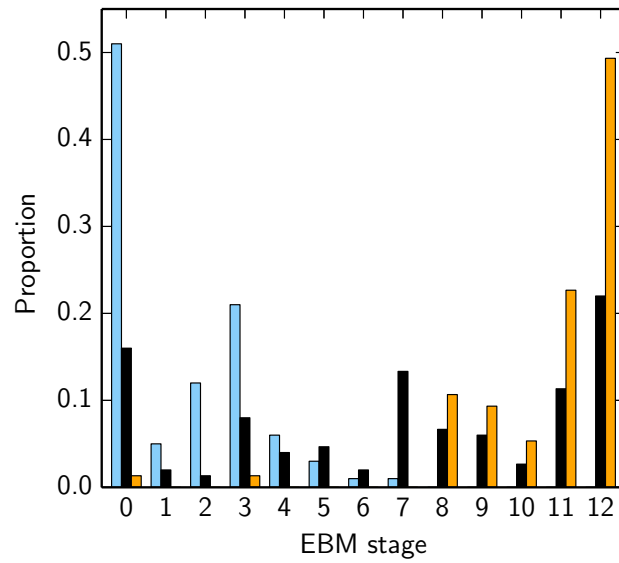
The balanced accuracy for predicting conversion (Table 3.5) was slightly reduced when the atrophy rates were removed but was still high, giving a maximum balanced accuracy of 71% (77% with atrophy rates) for conversion from MCI to AD over 3 years, and 70% (76% with atrophy rates) for conversion from CN to MCI over 5 years. On average over all follow up durations the balanced accuracy decreased by 2.6% for predicting conversion from MCI to AD, and increased by 4% for predicting conversion from CN to MCI. Again, increasing EBM stage was a significant hazard for conversion from both MCI to AD and CN to MCI (Table 3.6 and Figure 3.8).

	Demographics	CN	MCI	AD
All subjects	N	100	150	75
	Sex M/F	51/49 (51%)	98/52 (65%)	41/34 (55%)
	Age	75 $\pm$ 5	73 $\pm$ 7	75 $\pm$ 8
	Education	15.7 $\pm$ 2.9	15.7 $\pm$ 3	15.1 $\pm$ 3
	APOE4 +/-	22/78 (22%)	83/65 (57%)	52/23 (69%)
A $\beta$ +	N	36	111	69
	Sex M/F	20/16 (56%)	69/42 (62%)	38/31 (55%)
	Age	76 $\pm$ 5	74 $\pm$ 7	74 $\pm$ 8
	Education	15.9 $\pm$ 3.3	15.6 $\pm$ 3.1	15 $\pm$ 3
	APOE4 +/-	15/21 (42%)	74/37 (67%)	52/17 (75%)
APOE4+	N	22	85	52
	Sex M/F	15/7 (68%)	49/36 (58%)	30/22 (58%)
	Age	75 $\pm$ 6	73 $\pm$ 6	74 $\pm$ 8
	Education	15.6 $\pm$ 3.4	15.6 $\pm$ 3	14.8 $\pm$ 3
	APOE4 +/-	22/0 (100%)	85/0 (100%)	52/0 (100%)
A $\beta$ + APOE4+	N	15	74	52
	Sex M/F	10/5 (67%)	43/31 (58%)	30/22 (58%)
	Age	77 $\pm$ 6	74 $\pm$ 7	74 $\pm$ 8
	Education	15.5 $\pm$ 3.8	15.7 $\pm$ 2.9	14.8 $\pm$ 3
	APOE4 +/-	15/0 (100%)	74/0 (100%)	52/0 (100%)

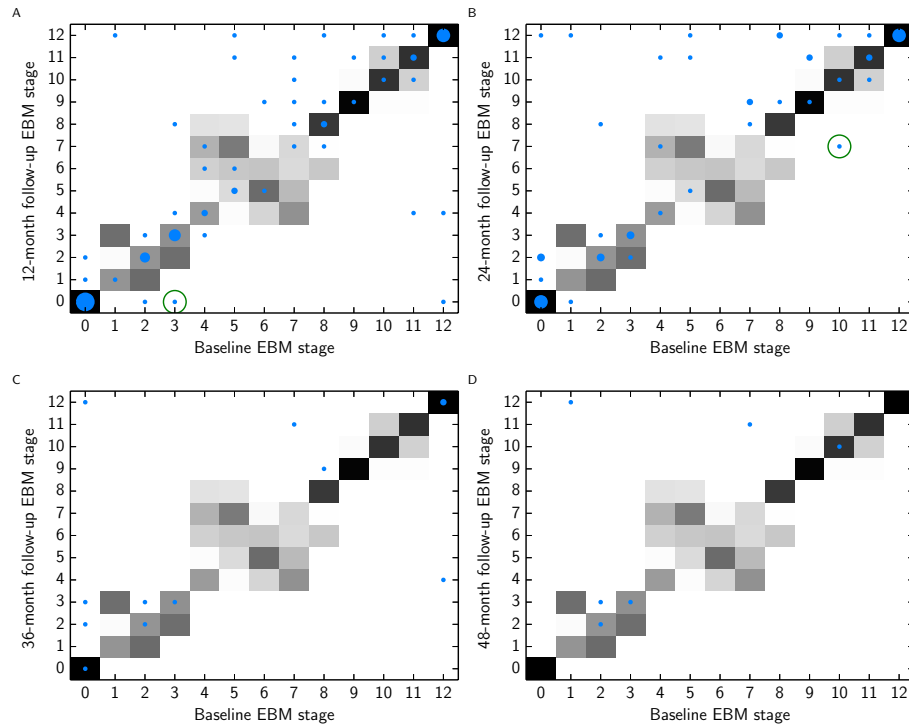
**Table 3.4:** As Table 3.1, but without using atrophy rates. Reproduced from [134]. CC BY 3.0.



**Figure 3.5:** As Figure 3.1, but without using atrophy rates, i.e. using the subjects in Table 3.4. Reproduced from [134]. CC BY 3.0.



**Figure 3.6:** As Figure 3.2, but without using atrophy rates. Events are ordered by the maximum likelihood event sequence for the whole population as shown in Figure 3.5. Reproduced from [134]. CC BY 3.0.



**Figure 3.7:** As Figure 3.3, but without using atrophy rates. Two additional follow up time points, at 36 and 48 months, met our inclusion criteria. Here, the largest dot, at (0,0) represents 17 subjects in (A) and 9 subjects in (B). The largest dot in (C) is at (12,12) and represents 2 subjects. In (D) all dots represent 1 subject. Reproduced from [134]. CC BY 3.0.

**A. MCI-stable vs. MCI-converters**

	Balanced Acc. (%)	Sensitivity (%)	Specificity (%)	AUC	Threshold Stage	N-c/N-s
12 months	67	72	62	0.71	8	32/103
24 months	68	68	68	0.71	8	57/68
36 months	71	83	59	0.74	6	69/51
48 months	74	84	65	0.71	5	74/20
60 months	73	84	63	0.74	5	77/16

**B. CN-stable vs. CN-converters**

	Balanced Acc. (%)	Sensitivity (%)	Specificity (%)	AUC	Threshold Stage	N-c/N-s
12 months	95	100	91	0.95	4	2/95
24 months	79	67	91	0.78	4	6/86
36 months	70	45	95	0.69	4	9/76
48 months	66	38	94	0.68	4	13/50
60 months	70	76	64	0.75	1	17/39

**Table 3.5:** As Table 3.2, but without using atrophy rates. Reproduced from [134]. CC BY 3.0.

### 3.5 Discussion

I have adapted the event-based model for use with multi-modal sporadic disease datasets to determine the characteristic ordering of biomarker transitions and provide a staging system for disease monitoring. I use the EBM here to derive characteristic biomarker orderings in AD from various subgroups of the ADNI dataset and to provide insight into the variability of the ordering. The orderings provide detailed information on the dynamics of large sets of biomarkers across the full duration of AD progression. They describe a distinct sequence of biomarker transitions in which CSF measures are the earliest to become abnormal, followed by atrophy rates, and finally cognitive test scores and regional brain volumes. The recovered ordering shows less variation in the sequence for  $A\beta+$ , APOE4+ or  $A\beta+$ APOE4+ individuals than for the whole population, most likely reflecting that the former are a more homogeneous group with archetypical AD pathology. The results of the EBM provide entirely data-driven support for hypothetical models of AD progression, such as [74, 78, 79], without the requirement for determining biomarker cut-points [108].



**A. MCI to AD progression**

	HR (CI)	p-value	Corrected HR (CI)	Corrected p-value
EBM Stage	1.16 (1.10-1.23)	$3.34 \times 10^{-7*}$	1.17 (1.10-1.24)	$3.55 \times 10^{-7*}$
Age	1.00 (0.97-1.03)	0.98	0.99 (0.96-1.02)	0.51
Education	0.98 (0.91-1.06)	0.65	0.98 (0.91-1.06)	0.6
APOE4 Carrier	1.56 (0.98-2.46)	0.059	1.32 (0.82-2.13)	0.25
Male	0.78 (0.49-1.22)	0.27	0.84 (0.50-1.43)	0.52

**B. CN to MCI progression**

	HR (CI)	p-value	Corrected HR (CI)	Corrected p-value
EBM Stage	1.66 (1.29-2.14)	$1.01 \times 10^{-4*}$	1.59 (1.22-2.09)	$6.72 \times 10^{-4*}$
Age	1.00 (0.91-1.10)	0.99	0.99 (0.90-1.09)	0.83
Education	1.02 (0.88-1.21)	0.76	0.99 (0.84-1.15)	0.85
APOE4 Carrier	3.00 (1.16-7.78)	0.024*	2.02 (0.68-6.00)	0.21
Male	2.00 (0.75-5.33)	0.17	1.38 (0.46-4.14)	0.57

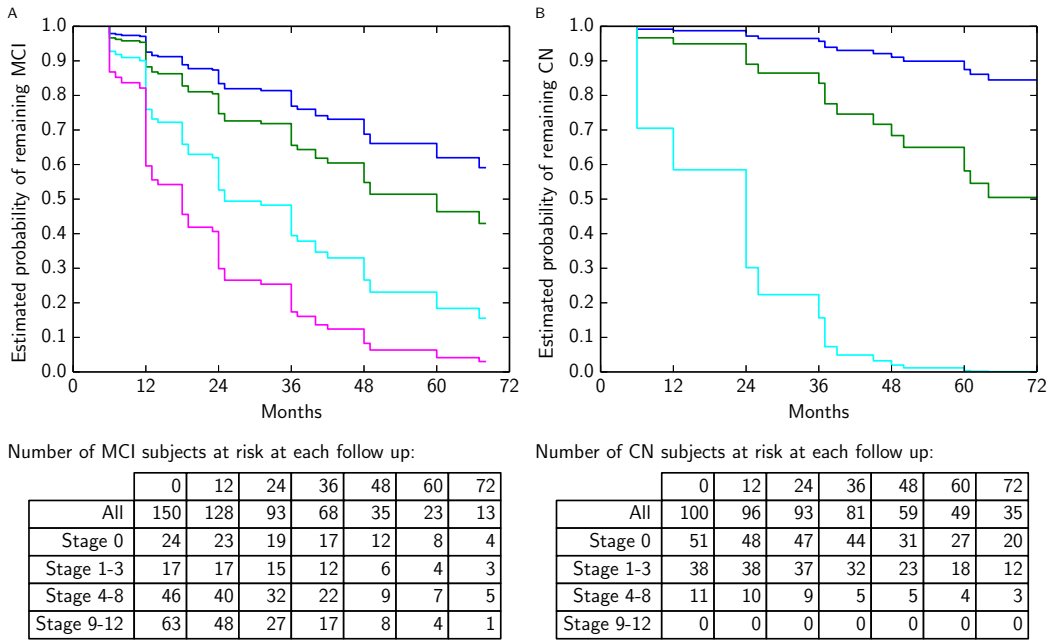
**Table 3.6:** As Table 3.3, but without using atrophy rates. Reproduced from [134]. CC BY 3.0.

The staging system provides a much more detailed evaluation of patient state than clinical diagnoses. Importantly, it has clear clinical relevance, providing a high accuracy classification of CN vs. AD subjects, predicting conversion from MCI to AD and CN to MCI, and being applicable not only to short-term longitudinal datasets (allowing atrophy measurements), but also to fully cross-sectional datasets (one visit).

### 3.5.1 Event sequence

#### 3.5.1.1 Ordering of CSF biomarkers

The ordering of the CSF biomarkers in  $A\beta+$  and APOE4+ individuals supports the ordering of CSF biomarkers predicted by earlier hypothetical models of AD progression: CSF  $A\beta_{1-42}$ , p-tau, t-tau. Since  $A\beta+$  individuals are likely to have early AD, this group should represent a much purer AD population than the whole population and thus the biomarker ordering should reflect the AD ordering more closely. Similarly, APOE4 carriers would also be predicted to show this pattern, given the very strong association between APOE4 and  $A\beta$  deposition [145].



**Figure 3.8:** As Figure 3.4, but without using atrophy rates. These estimated probabilities are shown for the average population demographics (74.2 years of age, 15.6 years of education, APOE4 negative, male sex). Stages are grouped analogously to Figure 3.4, so that here Normal (blue) = stage 0, CSF (green) = stages 1-3, Cognition (cyan) = stages 4-8, which includes hippocampal and entorhinal cortex volume as well as cognitive test scores, Volume (magenta) = stages 9-12. Reproduced from [134]. CC BY 3.0.

In the broader population, however, the results suggest that CSF t-tau and p-tau may become abnormal prior to  $A\beta_{1-42}$ , i.e. that there are a significant proportion of subjects who have CSF t-tau and p-tau but not  $A\beta_{1-42}$  abnormalities, although cross-validation shows higher uncertainty. Given the results in the APOE4+ and  $A\beta$ + populations, it seems likely that these subjects reside predominantly in the APOE4- and  $A\beta$ - populations, and indeed estimation of the ordering using the APOE4- and  $A\beta$ - subject groups alone supports this hypothesis, confirming that CSF t-tau and p-tau events appear earlier than CSF  $A\beta_{1-42}$  (data not shown). As discussed by Jack et al. [75, 146], there are several potential explanations for this finding. First, that tau accumulation is a common feature of aging. Braak et al. [147] found tau pathology to be present in healthy individuals at autopsy from as early as 20 years of age. These findings are replicated by the study of Kok et al. [148], which found NFT deposition in a significant proportion of APOE4 negative indi-

viduals between 30 and 59 years of age. These results, which demonstrate discrepancies between the ordering in APOE4 positive and negative individuals, would be entirely consistent with these findings, with the pattern observed in the population as a whole reflecting a mixture of two populations - one already on the path to developing AD, the other undergoing normal ageing, with t-tau and p-tau a common early feature in both. A second alternative is that accumulation of tau pathology may be a very early feature of AD either for some or all subjects. Early tau pathology may be more prevalent in APOE4- and A $\beta$ - individuals, or alternatively, as the subjects recruited for ADNI are age matched, early tau pathology might not be observed in the APOE4+ and A $\beta$ + populations who would be likely to develop AD at a younger age, and thus already have abnormal amyloid levels. A third possibility is that amyloid accumulation does precede tau deposition, but that either current CSF A $\beta$ <sub>1-42</sub> assays are less sensitive than the CSF t-tau and p-tau assays, or do not detect the very earliest (e.g. oligomeric) abnormal A $\beta$  moieties. Finally, as CSF t-tau is not specific to AD and is found in other neurodegenerative diseases, stroke, trauma and encephalitis [149], a further alternative is that individuals have other, perhaps presymptomatic neurodegenerative diseases, such as frontotemporal dementia (FTD), or dementia with Lewy bodies (DLB). Such individuals might be under-represented in the APOE4+ and/or A $\beta$ + groups, which are enriched for AD, and thus more prevalent in the APOE4- and A $\beta$ - groups.

### 3.5.1.2 Ordering of MRI biomarkers

The ordering of MRI biomarkers from the EBM agrees with previous findings (e.g. [113, 114]), with atrophy rates becoming abnormal prior to overall volume changes, and volume changes occurring in a distinct sequence, starting in the hippocampus and entorhinal cortex, progressing to other temporal lobe areas, the middle temporal gyrus and the fusiform gyrus, with resulting overall brain volume loss and ventricular expansion. Results in APOE4+ subjects also support previous findings [109, 150], suggesting earlier hippocampal and entorhinal volume loss, which occur prior to MMSE reduction in the APOE4+ population and after MMSE in the whole population and A $\beta$ + population.

One perhaps surprising result of the MRI biomarker ordering is that the increasing whole brain atrophy rate event occurs prior to the hippocampal atrophy rate event both in the whole and  $A\beta+$  population. In common with any data-driven model of biomarker changes, the EBM orders events based on when the corresponding measurements become discernibly different between cases and controls. This may not reflect the order of appearance of underlying pathology as the precision of the measurements may vary [98]. Thus, this result might simply reflect the increased variability associated with measurement of hippocampal over whole brain atrophy rates [151]. Other possible factors are that the results are influenced by subjects who have a mixture of pathologies, where other processes occur alongside AD which contribute to brain atrophy rate but not hippocampal atrophy rate, such as vascular disease [152], or other neurodegenerative diseases [153]. Alternatively, excess whole brain atrophy may be a core feature of all patients with AD, noting that some individuals with pathologically confirmed AD have relatively hippocampal sparing disease [154].

### 3.5.1.3 Uncertainty in the event sequence

The uncertainty in the event sequence, as shown by the positional variance diagrams and cross-validation results, potentially provides useful information about the variation of biomarker ordering across the population. However, three main factors contribute to the uncertainty. First, natural variation: some events may occur in different orders in different individuals. For example, for APOE4+ subjects, hippocampal volume loss may occur earlier than in APOE4 negative subjects [109, 150]; thus in the whole population that combines both groups, uncertainty is higher. Second, sampling density: when events occur in close succession, there are likely to be fewer of the data points, which are required to determine their ordering, that separate them. Third, outliers: the dataset may include subjects who do not follow any typical progression pattern of AD, e.g. subjects with other neurodegenerative diseases. Although the model fitting procedure I use is somewhat robust to these outliers, they can still affect the posterior distribution on the ordering, which manifests as uncertainty.

#### 3.5.1.4 Using the event-based model to define cut points

A major advantage of the EBM is that the ordering of biomarkers is not dependent on cut points. Instead, the EBM is probabilistic, calculating the probability that each event has occurred from models of the distributions of normal and abnormal biomarkers learned from the data rather than assuming an event has occurred when a certain threshold value is reached. However, for comparison I derived cut point values, given in Table 3.7, which represent the point at which the biomarker value is equally likely to be normal or abnormal, and should therefore be similar to existing biomarker cut points. The resulting cut points for the CSF biomarkers are very similar to those reported by Shaw et al. [139], which were derived using a maximum accuracy classification of autopsy confirmed AD vs. healthy controls. Importantly, the ordering provided by the EBM can be seen not merely to reflect the ordering of the sensitivity or specificity of these cut points.

### 3.5.2 Patient staging

A more directly practical output of the EBM is the data-driven staging system it provides. Here I demonstrate, for the first time, the use of such a patient staging measure to predict clinical outcomes. The EBM's staging measure strongly separates CN and AD subjects and gives comparable results to state of the art classification techniques for prediction of conversion from MCI to AD [121] albeit with a larger set of biomarkers. The major advantage of the EBM, a generative model, is that it explicitly provides useful information on what drives the classification unlike the discriminative models used in [121]. I used the EBM's staging system to predict conversion from CN to MCI, as well as MCI to AD, and over different follow up durations. The classification results are supported by the results of the Cox proportional hazards models, which find EBM stage to be a significant hazard for conversion from both MCI to AD and CN to MCI. This suggests that the EBM, once sufficient control/AD data are available, might have clinical application, providing valuable prognostic information on an individual patient basis, and potentially for clinical trial stratification.

Biomarker	Cut Point	Sensitivity (%)	Specificity (%)
Abeta (pg/ml)	189	92	63
P-tau (pg/ml)	25	88	71
T-tau (pg/ml)	80	77	73
Hippo. Atrophy (ml/year)	0.138	72	75
Brain Atrophy (ml/year)	11.9	64	78
RAVLT	33	92	91
ADAS-Cog	17	97	97
MMSE	27	100	97
Hippocampus (% TIV)	0.423	81	82
Entorhinal (% TIV)	0.214	84	83
Mid Temporal (% TIV)	1.19	75	78
Whole Brain (% TIV)	64.6	73	66
Fusiform (% TIV)	1.05	73	73
Ventricles (% TIV)	3.04	48	85

**Table 3.7:** Cut point values derived using the event distributions estimated by the EBM. Volume measurements (hippocampus, entorhinal, mid temporal, fusiform, whole brain, ventricles) are summed over the left and right hemisphere and total intracranial volume normalized, and are recorded as a percentage of the total intracranial volume. The sensitivity is the percentage of Alzheimers disease subjects with abnormal measurements, and specificity is the percentage of cognitively normal subjects with normal measurements, when subjects are classified using these cut points. MMSE = Mini-Mental State Examination; RAVLT = Rey Auditory Verbal Learning Test; TIV = total intracranial volume. Reproduced from [134]. CC BY 3.0.

### 3.5.3 Model assumptions

When interpreting these results, it is important to stress that the EBM is based on strong assumptions, which are explicitly designed to simplify reality in order to determine major trends in data. This section summarises the key assumptions made in the modelling process, their potential influence on results, and possibilities to relax the assumptions in future work.

#### 3.5.3.1 Event Sequence

The EBM, like other data-driven models [49, 80, 81, 83, 104–106], assumes that all subjects follow a single progression pattern. While this may be reasonable for the  $A\beta+$  and APOE4+ groups, the wider sporadic AD is likely to show more variability in the event sequence due to the inherent disease heterogeneity, driven perhaps by genetic, e.g. the presence or absence of APOE4 [155], or lifestyle factors. The sin-

gle sequence the EBM identifies maximises compatibility within the set of subjects. It is thus important to consider not only the most likely sequence, but also the positional variance diagram and cross-validation output, which explicitly highlight areas of uncertainty, aiding interpretation particularly where the data departs from the assumptions, for example in heterogeneous groups. The positional variance diagrams generated directly from the EBM (Figure 3.1 A-D) underestimate the uncertainty in the event ordering, as they do not account for uncertainty in the biomarker distribution models. The cross-validation results (Figure 3.1 E-H), on the other hand, tend to overestimate the uncertainty, because each iteration considers only a subset of the data. In the whole-population analysis, both mechanisms show reasonable stability of the results, which gives some confidence to the conclusions. However, it is important to remember that the single sequence does not represent all subjects and the positional variance diagrams are only a crude indicator of heterogeneity of the event sequence. More sophisticated models that can relax the assumption of a single event ordering, see for example [156, 157], and/or provide uncertainty estimates by modelling the uncertainty in the biomarker distribution parameters, are important areas for future study.

### 3.5.3.2 Patient Staging

Whilst the modelling approach provides a powerful potential means of patient staging it is important that such staging information is interpreted correctly. While the idealised model for, say, stage 3 is that all CSF biomarkers are abnormal and all others are normal, a patient assigned stage 3 need not fit this profile exactly; stage 3 is simply the idealised stage most compatible with a given individual's biomarker measurements. This formulation enables the EBM to stage subjects who do not conform to the maximum likelihood event sequence, which is important given the heterogeneity of sporadic AD. Despite its idealised nature, the staging system has clear clinical relevance, as demonstrated by the strong classification performance and Cox proportional hazards Model results; those results also add confidence to the event sequence derived from the whole population, which underpins the staging. The probabilistic nature of the staging system presents opportunities for refinement

in future work. Here I assign only the most likely stage, but using equation (3.2) I can quantify the uncertainty in the stage assignment, which may contain useful additional diagnostic and prognostic information. Moreover, also using equation (3.2), I can obtain an overall likelihood of conforming to the event sequence, which should be useful for detecting misdiagnoses or choosing the most likely diagnosis from a selection of models for different diseases.

## 3.6 Conclusion

I have developed a data-driven model for determining biomarker ordering in sporadic AD (contribution 1.3.1 (a), Chapter 1, page 39). I have used the model with the ADNI data set to support currently hypothetical models (contribution 1.3.1 (b), Chapter 1, page 39), but further to highlight uncertainty in those orderings and variation among different subgroups. I also demonstrate that such a model can provide a practical and effective staging system for patient prognosis.



## Chapter 4

# A data-driven model of biomarker changes in dominantly-inherited Alzheimer's disease

This chapter details sub-contributions (c) and (d) of contribution 1.3.1 (see Chapter 1, page 39). The work presented in this chapter forms part of a collaboration led by Neil Oxtoby to look at biomarker changes in dominantly-inherited AD using two types of data-driven model: event-based models and differential equation models. We analysed data from the Dominantly Inherited Alzheimer Network (DIAN): I performed the event-based model analyses, the results of which are presented in this chapter; Neil performed the differential equation model analyses (not included here). I wrote the initial drafts of the event-based model sections of the manuscript, but Neil and our co-authors contributed to the refinement of the text. This work was presented at the *Alzheimer's Association International Conference* in 2016 and is currently in preparation for publication.

### 4.1 Associated publications

N. P. Oxtoby, A. L. Young, D. M. Cash, T. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, and D. C. Alexander. Data-driven models for predicting fine-grained disease progression and symptom onset in dominantly-inherited Alzheimer's disease without reliance upon familial age of onset. *In Prepa-*

## 4.2 Introduction

Dominantly-inherited Alzheimer's disease (DIAD) is a rare early onset form of AD in which subjects can be identified presymptomatically by the presence of a pathogenic mutation in one of three genes: amyloid precursor protein (APP), presenillin 1 (PSEN1) or presenillin 2 (PSEN2). As such, it provides a window into the presymptomatic phase of AD, which is of great interest as it is likely that treatments will be most efficacious if given early.

Quantitative models of AD biomarker changes can provide a detailed picture of disease progression and a potential mechanism to identify and stage patients presymptomatically. Previous studies of biomarker changes in DIAD consider biomarker values as a function of estimated years from onset (EYO) of clinical symptoms (e.g. [49]), which can be estimated using parental age of onset, or mean family age of onset. There is considerable uncertainty in this estimate of familial age of onset because (1) it is difficult to ascertain when an individual became affected, and (2) there can be substantial differences in actual age of onset within families and mutations. This uncertainty limits the utility of EYO for estimating disease progression in presymptomatic DIAD individuals, reducing the resolution at which biomarker ordering can be determined.

This work uses a data-driven model of disease progression to explore the sequence of biomarker changes in DIAD without reliance upon EYO. I extend the event-based model (EBM) for use with missing data, facilitating its application to multi-modal DIAD data from the Dominantly Inherited Alzheimer Network (DIAN). I further explore the utility of this model for patient staging.

## 4.3 Materials and methods

### 4.3.1 Data description

For detailed descriptive summaries of the DIAN cohort see [159]. At the sixth data freeze, the DIAN cohort included 338 individuals with known mutation status and

a baseline visit, aged 19-66 at baseline with up to four visits each ( $1.1 \pm 1.9$  years in duration), spanning up to 30 years before and up to 21 years after parental age of symptom onset. Of these, 211 individuals carry DIAD mutations: 163 PSEN1; 17 PSEN2; and 31 APP. Mutation carriers were further subdivided into diagnostic groups: ‘Cognitively Normal’ (CN), ‘Mild Cognitive Impairment’ (MCI), and ‘Alzheimer’s Disease’ (AD). The definition of these diagnostic groups is slightly different to that of the ADNI cohort in Chapters 3, 5, 6, and 7. Here the definition is based on the Clinical Dementia Rating (CDR) scale alone: CN is defined as a global CDR of 0, MCI is defined as a global CDR of 0.5, AD is defined as a global CDR of  $\geq 1$ . In ADNI the definition is based not only on the CDR, but also on the Mini Mental State Examination (MMSE) score, and various other clinical criteria<sup>1</sup>.

AD biomarkers were selected based on specificity to the disease, or if disease ‘signal’ is present, i.e. quantifiable distinction between mutation carriers and non-carriers. The biomarkers include CSF measures of molecular pathology (amyloid proteins and neurofibrillary tangles); a cognitive test score (MMSE); regional brain volumetry from MRI, e.g. hippocampus, middle-temporal region, temporo-parietal cortex; PiB-PET imaging measures of amyloid accumulation; and FDG-PET imaging measures of glucose hypometabolism. The full set of biomarkers included in the EBM is listed on the vertical axis of Figure 4.1. Eight biomarkers (regional MRI volumes of the caudate, entorhinal cortex, pallidum and thalamus, and FDG standard uptake value ratio (SUVR) measurements in the accumbens, caudate, putamen and thalamus) were excluded from the EBM because there were no statistically significant differences (Bonferroni corrected p-value of  $p < \frac{0.01}{N}$ , where  $N$  - the total number of biomarkers considered for inclusion in the EBM - equals 29) on a two-sample t-test comparing the sample mean of the non-carriers with the sample mean of the mutation carriers that have an AD diagnosis. Baseline data for mutation carriers and non-carriers was used to fit EBMs.

I used stepwise regression to remove the influence of age, years of education, sex, and head size (total intracranial volume) prior to fitting the models. Individuals

---

<sup>1</sup>For details see [www.adni-info.org/Scientists/Pdfs/adniproceduresmanual12.pdf](http://www.adni-info.org/Scientists/Pdfs/adniproceduresmanual12.pdf).

with missing genetic, gender, and education data were excluded. I also excluded 21 structural MRI scans from 10 participants due to image artefacts or non-AD pathology. Table 4.1 summarises the included participant data and their demographics.

Stepwise regression is a simple statistical method for performing linear regression when it is not known beforehand if there is a statistical relationship between the predictor variables and the measured data. Common criticisms of stepwise regression are that it is used incorrectly - in cases where there are many predictor variables but not enough data to estimate the coefficients meaningfully, and that it may not provide the globally optimal model. Here, there are only a few predictor variables (age, years of education, sex, and head size), so there is enough data to estimate meaningful coefficients if there is a relationship between the predictor and the measured data, and a locally optimal model should be sufficient to broadly remove the influence of a particular variable. Using stepwise regression means that the predictor variables are only corrected for when they reach statistical significance. Although a lack of statistical significance does not necessarily mean there is no relationship between the predictor variable and the measured data, only correcting for statistically significant predictors does prevent unnecessary confounds from being introduced, as well as avoiding the correction model becoming overly complex. A complex correction model reduces the interpretability of the corrected values of the measured data (biomarkers).

Demographic	NC	MC [PSEN1, PSEN2, APP]
n analysed	127	211 [163 (77%), 17 (8%), 31 (15%)]
Female (%)	75 (59%)	117 (55%) [92 (79%), 5 (4%), 20 (17%)]
ApoE4-positive (%)	37 (29%)	61 (29%) [47 (77%), 7 (11.5%), 7 (11.5%)]
ApoE4-negative (%)	90 (71%)	150 (71%) [116 (77%), 10 (7%), 24 (16%)]
Baseline Age (SD)	39 (10)	39 (10) [39 (10), 39 (10), 43 (10)]
Baseline Education (SD)	15 (3)	14 (3) [14 (3), 15 (3), 14 (3)]
Baseline EYO (SD)	-7 (12)	-7 (10) [-7 (10), -12 (10), -6 (9)]

**Table 4.1:** Demographics for DIAN participants having cross-sectional data at Data Freeze 6, used to build Event-Based Models of DIAD. Age, education and EYO measured in years. EYO = estimated years from onset; SD = standard deviation; NC = non-carriers; MC = mutation carriers.

### 4.3.2 Event-based models

I fitted an EBM to determine the most probable sequence of biomarker abnormality events and the uncertainty in this sequence [98]. Each event represents the transition of a biomarker from a normal (as seen in non-carriers) to an abnormal level. The probability a biomarker measurement is normal is modelled as a gaussian distribution, and estimated using data from non-carriers. The distribution of abnormal measurements is also modelled as a gaussian distribution, but estimated by fitting a mixture of two gaussians [98] to data from all mutation carriers: the first gaussian models the distribution of normal measurements, and is kept fixed to the values estimated from non-carriers; the second gaussian models the distribution of abnormal measurements, and is optimised using data from mutation carriers.

The sequence of events was estimated in various population subgroups: all 191 mutation carriers; 150 PSEN1 mutation carriers; 14 PSEN2 mutation carriers; and 26 APP mutation carriers. I also considered separate EBM models by APOE4 status: 58 mutation carriers who were APOE4-positive (with one or more APOE4 alleles), and 133 mutation carriers who were APOE4-negative. I accounted for missing data by imputing biomarker values such that missing measurements had an equal probability of being normal or abnormal, and thus do not influence the population sequence. I found the value for imputation by simply computing the points at which the distribution of normal and abnormal measurements intersect, and choosing the value that was between the means of the two distributions. The uncertainty in the event sequence was estimated by taking Markov Chain Monte Carlo (MCMC) samples from the EBM [98]. Cross-validation of the event sequence was performed by refitting the biomarker event distributions and the event sequence for 100 bootstrap samples from each data subset.

I assigned subjects to patient stages based on their most probable position along the most probable event sequence [134] for all mutation carriers combined. I assessed the efficacy of the patient staging system using only subjects with data available for all biomarkers ( $n = 30$ , total of 42 followup visits), as missing entries cause uncertainty in a subject's model stage.

### 4.3.3 Implementation

All experiments were performed using Matlab on a standard workstation (Intel Core i7, 3.1 GHz, 8GB memory). Stepwise regression was implemented using the Matlab 'stepwiselm' algorithm. The event distribution (gaussian mixture model) parameters were optimised using the 'fmincon' constrained optimisation solver with the 'sqp' - sequential quadratic programming - method. All of the gaussian mixture model fits were assessed visually to ensure the solutions were physically plausible. To find the characteristic event sequence  $\bar{S}$ , I performed 2,500 iterations of a greedy ascent algorithm (as in [98]), initialised from 25 random start points. I checked that the start points converged to a single maximum to ensure the global optimum had been found. I then ran an MCMC algorithm (as in [98]) to draw samples from the posterior distribution  $P(S|X)$ . I ran the MCMC algorithm for 100,000 iterations, checking that the MCMC trace showed good mixing properties. I initialised the MCMC algorithm at  $\bar{S}$  so a burn-in period was not required. Computational complexity is as in Chapter 3.

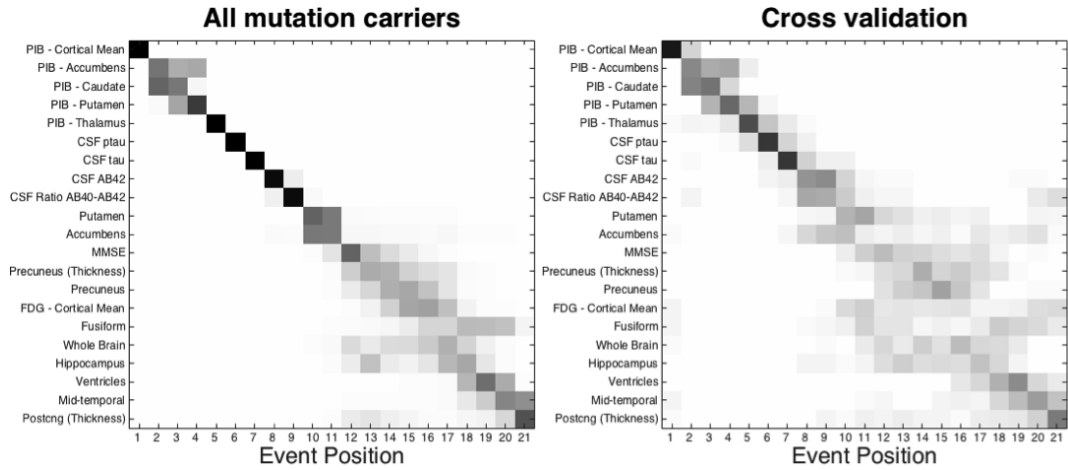
## 4.4 Results

### 4.4.1 Event sequences

#### 4.4.1.1 All mutation carriers

Figure 4.1 is a positional variance diagram of the maximum likelihood sequence of biomarker abnormality events (top to bottom), and its uncertainty (left to right), across all available 211 mutation carriers in the DIAN dataset. Grayscale intensity represents confidence in each events position within the sequence, and is calculated from MCMC samples from the EBM [134].

The EBM reveals a distinct sequence of biomarker abnormality in DIAD: regional (cortical then striatal) amyloid deposition on PiB-PET scans; CSF measures of neuronal injury/neurofibrillary tangles (total tau and phosphorylated tau levels) and amyloid plaques ( $A\beta_{42}$  and  $A\beta_{40/42}$  ratio); MR measures of volume loss in the putamen and nucleus accumbens; global cognition (MMSE score). Thereafter the ordering in which FDG-PET hypometabolism and other MRI measures



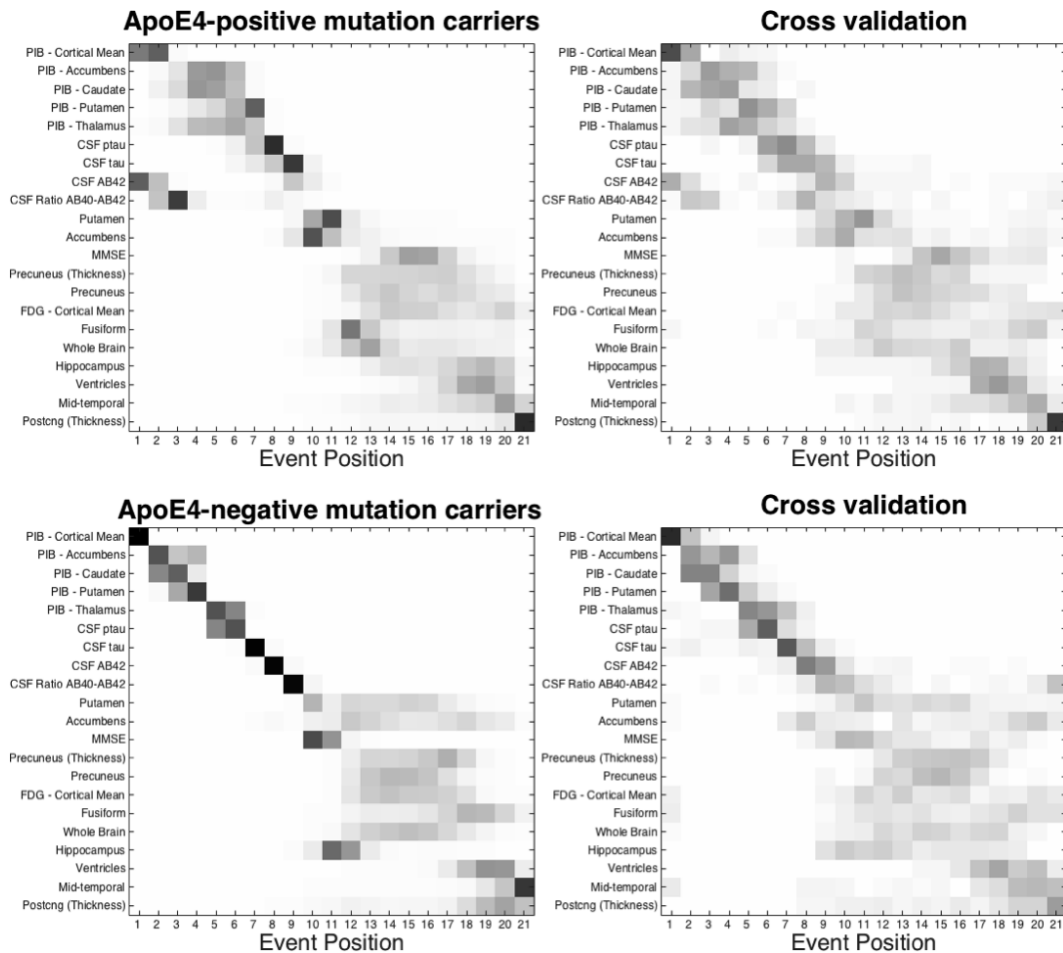
**Figure 4.1:** EBM progression of DIAD represented as positional variance diagrams. Left: EBM estimated on all mutation carriers in the DIAN dataset. Right: cross-validation through bootstrapping. The vertical ordering (top to bottom) is given by the maximum likelihood sequence estimated by the model. Grayscale intensity (left to right) represents MCMC-sampled posterior confidence in each events position.

become abnormal is less certain. The ordering shows high certainty early in the ordering of these biomarkers (as reflected by the more solid blocks along the diagonal), with lower certainty later in the ordering of regional volumes (more diffuse grey blocks straying from the diagonal). This pattern (left) persists under cross-validation (right).

#### 4.4.1.2 Population subgroups

I also fit the EBM to subgroups of the mutation carriers in the DIAN data set. Figure 4.2 shows positional variance diagrams of the biomarker abnormality event sequence in APOE4-positive and APOE4-negative participants (those with and without the apolipoprotein- $\epsilon$ 4 allele), and Figure 4.3 shows equivalent results for the three DIAD mutation types in DIAN: PSEN1, PSEN2, and APP. For ease of comparison, the sequence ordering on the vertical axes of each plot is chosen to be the most probable ordering from Figure 4.1 (all mutation carriers).

Broadly speaking, there is good agreement of the event sequences across subgroups in Figures 4.2 and 4.3, with some subtle differences between groups: earlier CSF A $\beta$ 42 and A $\beta$ 40/42 ratio in the APOE4-positive and APP groups; earlier fusiform volume for the PSEN2 group; earlier putamen volume abnormality for the



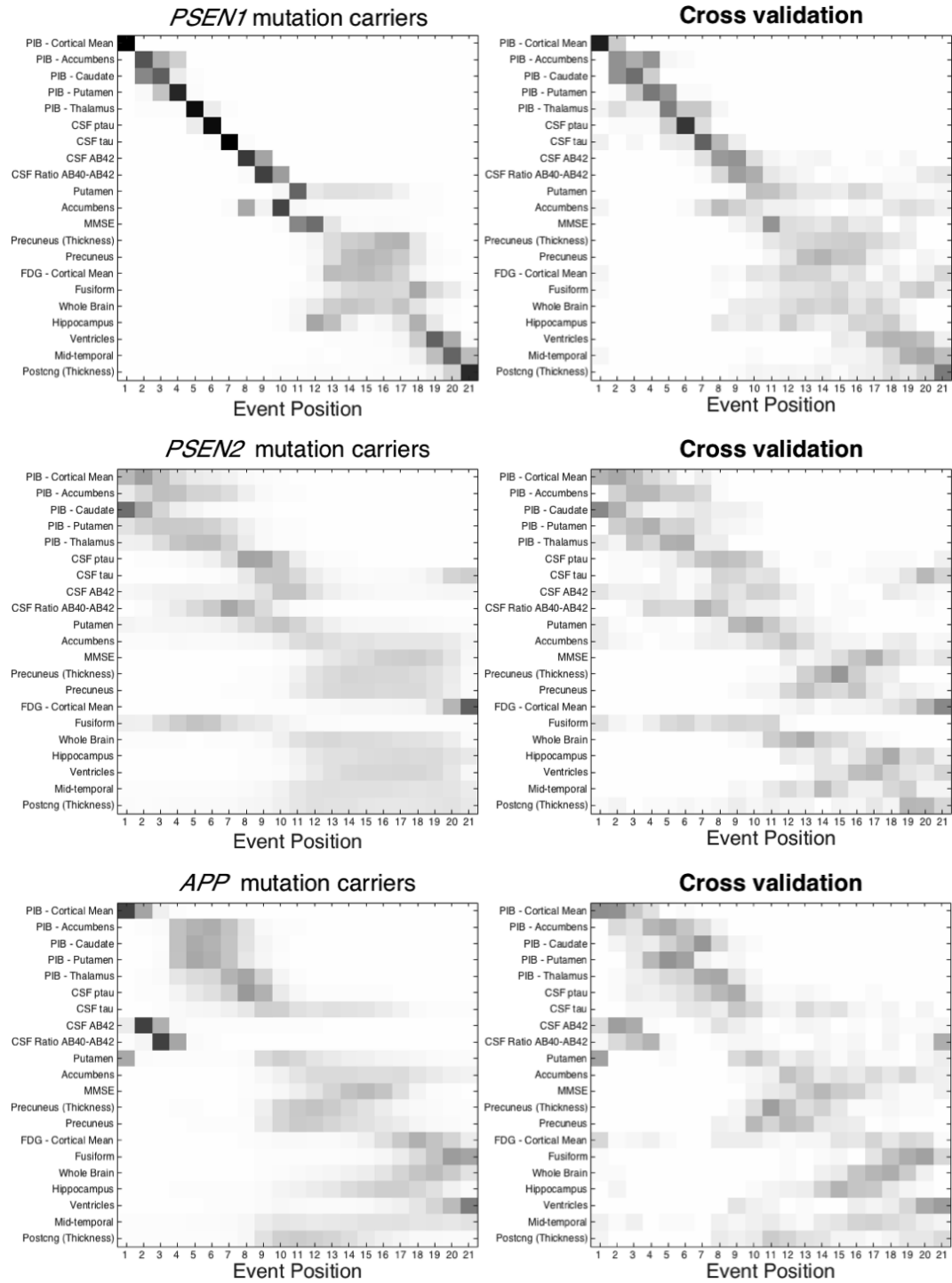
**Figure 4.2:** As figure 4.1, but for APOE4-positive and APOE4-negative mutation carriers.

APP group. The uncertainty is high in the subgroup orderings, due in part to the low numbers of participants in these groups, which reduces power to draw concrete conclusions based on these subtle differences between groups.

#### 4.4.2 Patient staging

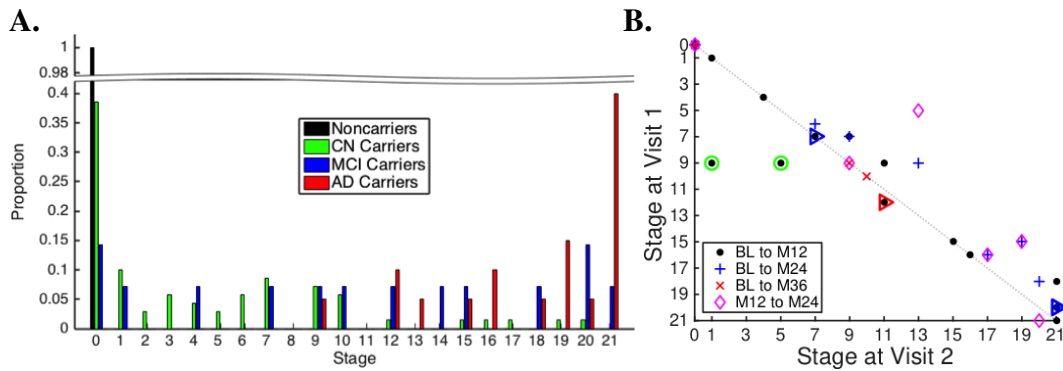
Figure 4.4 demonstrates the fine-grained staging capabilities of the EBM. Using the model for all mutation types (Figure 4.1), each individual in the DIAN dataset was assigned a disease stage that best reflects their measurements (see Methods section and [134]). The staging proportions are shown in Figure 4.4 A, differentiated by broad diagnostic groups (CN: Cognitively Normal, global CDR = 0; MCI: Mild Cognitive Impairment, global CDR = 0.5; AD: probable dementia due to AD, global CDR  $\geq 1$ ). Longitudinal consistency of staging is shown in Figure 4.4 B, where each individual's baseline stage is plotted against available follow-up stages





**Figure 4.3:** As figure 4.1, but for *PSEN1*, *PSEN2*, and *APP* mutation carriers.

between baseline and months 12/24/36.



**Figure 4.4:** DIAD: EBM staging results. A. Proportions grouped by diagnostic group: all noncarriers are at stage zero (black), and advancing disease stage is correlated strongly with cognitive impairment (green to blue to red). B. Staging consistency across visits within three years of baseline for the  $n = 30$  participants with complete longitudinal data (18 MCs; 16 PSEN1, 2 APP). Most individuals advance to a later stage (disease progresses towards the right). Green circles show the two individuals who regressed to earlier EBM stages, which arises in both cases due to discordant amyloid measurements between CSF and PiB-PET. The triangles indicate clinical progressors (CN to MCI in blue; MCI to AD in red). BL: baseline; M: month; CN: Cognitively Normal (global CDR = 0); MCI: Mild Cognitive Impairment (global CDR = 0.5); AD: probable dementia due to DIAD (global CDR  $\geq 1$ ).

#### 4.4.2.1 Cross-sectional distribution of stages

The baseline staging in Figure 4.4 A shows good separation of diagnostic groups: all of the non-carriers are assigned to stage 0 (black), CN mutation carriers (green) are at earlier model stages, mutation carriers diagnosed with probable AD dementia (red) are at late model stages, and mutation carriers diagnosed with MCI (blue) are more spread out across the stages. Within carriers, the model shows high classification accuracy for separating those who are CN from those with probable dementia: a balanced accuracy of 93% is achieved by classifying participants above stage 10 (putamen volume abnormality) as having probable dementia.

#### 4.4.2.2 Longitudinal consistency

The follow-up staging in Figure 4.4 B shows good longitudinal consistency: at 32 of 36 (89%) follow-up time points the model stage is the same or it increased; at 34 of 36 (94%) follow-up time points the stage was either unchanged, it increased, or it decreased within the uncertainty of the ordering. This included the clinical converters, which are shown with triangles (CN to MCI in green; MCI to AD in

red). The remaining two follow-up time points at which the model stage decreased (green circles in Figure 4.4 B) have inconsistent amyloid levels between CSF and regional PiB-PET, potentially due to discord between these biomarkers as has been observed in some individuals [160, 161].

## 4.5 Discussion

I have reported data-driven estimates of DIAD progression using an event-based modelling approach without reliance upon the use of familial age of onset as a measure of disease progression. The model reveals the sequence in which biomarkers of DIAD become abnormal. The resulting sequence broadly agrees with current understanding of DIAD, but provides superior detail compared to previous studies.

### 4.5.1 Event sequences

The EBM finds a distinct ordering of biomarker abnormality events in mutation carriers: amyloid deposition measured by PiB-PET, neurofibrillary tangles and amyloid plaques in CSF, followed by an AD-characteristic pattern of regional volume loss on MRI, which is interspersed with declining cognitive test scores and hypometabolism measured by FDG-PET. Although the sequence shows strong agreement across different mutation types (APP, PSEN1, PSEN2), and APOE4 positivity and negativity, I found some small, subtle differences that warrant further investigation. For example, there was earlier abnormality in CSF A $\beta$ 42 (than CSF tau) in the APP and APOE4-positive groups, but the reverse was found in other groups. The latter could be explained by non-monotonic dynamics of CSF A $\beta$ 42 in DIAD (an increase followed by a decrease) as suggested by results in previous investigations [162, 163]. Previous multimodal biomarker studies of DIAD, e.g. [49, 164, 165], are in general agreement with the EBM sequence: amyloidosis precedes hypometabolism and atrophy. Importantly, all previous approaches relied upon a familial age of symptom onset as a proxy for disease progression, which intrinsically limits the accuracy of predictions due to the known imprecision in such estimates [166].

### **4.5.2 Patient staging**

The staging system provided by the EBM has potential practical utility. In particular, it provides high classification accuracy for discriminating between presymptomatic and symptomatic DIAD mutation carriers; it correctly assigned all non-carriers to the ‘completely normal’ category (stage 0); and it showed good longitudinal consistency, with EBM stage generally increasing or remaining stable at patient follow-up. Thus it may be useful for identifying and monitoring presymptomatic individuals for clinical trials.

The EBM stages correlate strongly with cognitive status: CN participants were assigned early model stages, symptomatic AD participants late model stages, and MCI participants were more spread out across the stages. The MCI group in DIAD were the most heterogeneous, which is in agreement with the EBM results for sporadic AD [134], but possibly for different reasons. One contributing factor in DIAD is that the MCI group may include unaffected mutation carriers whose anxiety about their mutation status manifested as apparent cognitive abnormality and contributed to a diagnosis of MCI. In any case, the fine-grained disease staging offered by the EBM can shed light upon the heterogeneity contained within the MCI diagnostic stage.

## **4.6 Conclusion**

I have proposed an extension of the EBM to account for missing data (contribution 1.3.1 (c), Chapter 1, page 39), facilitating the application of the EBM in DIAD to determine the sequence of biomarker changes without reliance upon EYO (contribution 1.3.1 (d), Chapter 1, page 39). I have reported estimates of this sequence in various population subgroups, and demonstrated the ability of the EBM to provide fine-grained patient staging.

## Chapter 5

# A simulation system for biomarker evolution in neurodegenerative disease

This chapter details contribution 1.3.2 (see Chapter 1, page 40). The work presented in this chapter was published in *Medical Image Analysis* in 2015. I developed the methodology, performed the analysis and wrote the manuscript; my co-authors provided feedback on the methodology and the manuscript.

### 5.1 Associated publications

A. L. Young, N. P. Oxtoby, S. Ourselin, J. M. Schott, and D. C. Alexander. A simulation system for biomarker evolution in neurodegenerative disease. *Medical Image Analysis*, 26(1):47–56, 2015

### 5.2 Introduction

Data-driven models of disease progression allow longitudinal trends to be reconstructed from cross-sectional or short-term longitudinal datasets. Basic techniques to analyse biomarker trajectories involve staging subjects and then comparing biomarker levels across different disease stages [49, 83, 84, 104–106, 109, 110, 168]. This limits the temporal resolution of the model to the accuracy of the patient staging. Data-driven models do not require prior knowledge of the stage of a patient

along the disease, allowing the reconstruction of a much more fine-grained picture of disease progression. Differential equation models (DEMs) [81, 82, 110, 131] model short-term longitudinal patient data as a differential cross-section of a common longitudinal biomarker trajectory; the overall biomarker trajectory is obtained by integrating all of the subject's differential cross-sections. Jedynak et al. [132] and Donohue et al. [128] make a similar set of assumptions to formulate their disease progression models, again modelling each subject's biomarker data as a snapshot of a common progression curve, but further allowing for variation in individual progression rates. Another data-driven model, the event-based model (EBM) [98, 134] considers disease progression as a sequence of events at which biomarkers become abnormal, thereby allowing direct determination of biomarker ordering from entirely cross-sectional data.

However, such data-driven models typically depend on idealised assumptions about the data that they are modelling. First, that all subjects follow the same progression pattern. This is not true in general as large cross-sectional datasets will contain subjects who have different disease subtypes, mixed pathology, have been misdiagnosed, are yet to develop other diseases, or who are aging healthily. Such outliers are particularly prevalent in presymptomatic populations where the diagnostic outcome is unknown. Second, a set of parameters that define normal and abnormal biomarker levels. This is difficult to determine due to the high proportions of presymptomatic subjects in typical control populations (for example, a significant proportion of cognitively normal (CN) elderly subjects have been found to have biomarker changes consistent with AD [135, 136]), and misdiagnosis in diseased populations. Third, that the underlying disease time course is well sampled. In reality, presymptomatic subjects may not go on to develop the neurodegenerative disease being investigated and therefore the early disease stages might be under sampled or misrepresented, and diseased subjects may not be representative of the very late disease stages where the population thins and severe illness can make data hard to collect.

Here I present a simulation system to generate synthetic biomarker datasets

that represent the heterogeneity of sporadic neurodegenerative diseases. Although still based on a model of disease progression, it encapsulates many more variables than the simpler models that are parsimonious enough to fit to current data sets. Thus it provides a platform to evaluate the effect of more brutal simplifications necessary to obtain robust fitting results from working models. I demonstrate this simulation system by evaluating the performance of the EBM and a DEM in determining the sequence of biomarker abnormality from simulated data.

## 5.3 Materials and methods

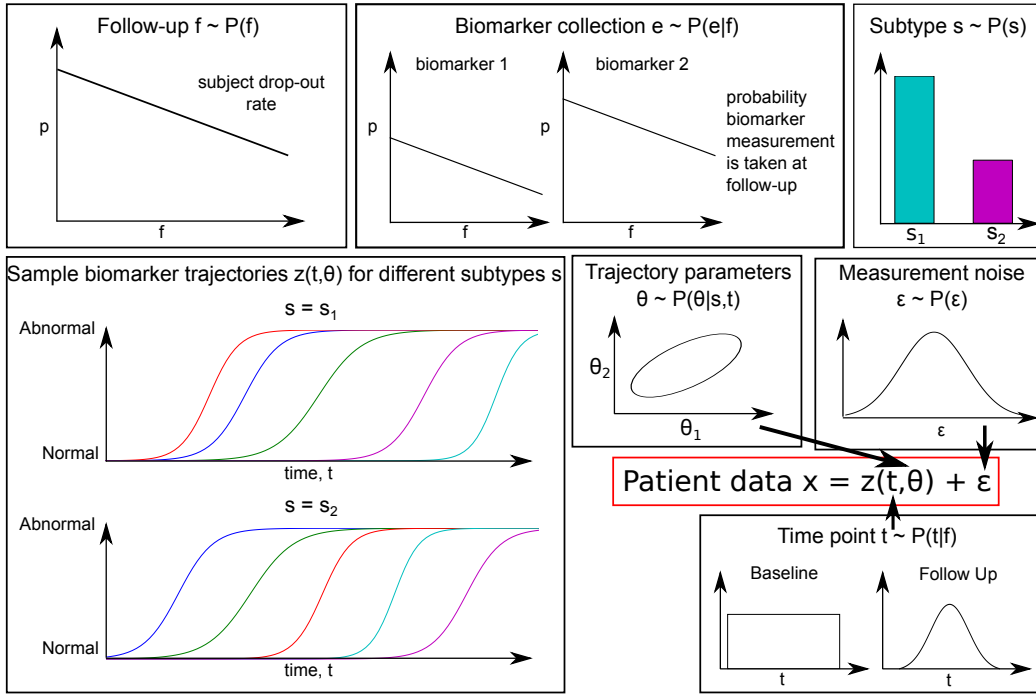
### 5.3.1 Generative model of data

I assume the following generative model of sporadic disease datasets (Figure 5.1). A set of subjects with: follow-up time points,  $\vec{f}$ , from a follow-up distribution  $P(\vec{f})$ ; disease subtypes,  $s$ , from a subtype distribution  $P(s)$ ; each set of follow-up time points correspond to a set of time points along the disease,  $\vec{t}$ , from a time point distribution  $P(\vec{t}|\vec{f})$ . At each time point a subset,  $\vec{e}$ , of the biomarkers included in the study are measured for a particular subject according to a biomarker collection distribution  $P(\vec{e}|\vec{f})$ , i.e. a subset of the biomarkers included in the study are measured in the subset of subjects that have a particular follow-up visit. Each subject has a set of biomarker measurements,  $\vec{x}_{\vec{e}}$ , at each time point  $t \in \vec{t}$ . The collected biomarker measurements are simulated from a trajectory evolution function  $z(t, \vec{\theta})$  with parameters  $\vec{\theta}$  from a trajectory parameter distribution  $P(\vec{\theta}|s, t)$ , and measurement noise  $\epsilon$  perturbation from a measurement noise distribution  $P(\epsilon)$ . The data for each collected biomarker for each patient for each time point is then  $x = z(t, \vec{\theta}) + \epsilon$ ; biomarkers that are not collected are recorded as missing data. Each subject is given a particular diagnosis  $d$  from a diagnosis distribution  $P(d|\vec{x}_{\vec{e}})$ .

### 5.3.2 Simulating sporadic Alzheimer's disease

#### 5.3.2.1 ADNI dataset

I downloaded baseline and follow-up data from all subjects in ADNI-1 giving a set of 819 subjects, (229 CN, 398 mild cognitive impairment (MCI), 192 AD). I



**Figure 5.1:** Generative model of sporadic disease datasets used by the simulation system. Reproduced from [167]. CC BY 4.0.

included the following set of biomarker data: CSF  $A\beta_{1-42}$ , CSF t-tau, CSF p-tau; the Mini Mental State Examination (MMSE) [85]; baseline MRI volumes of whole brain, hippocampus and ventricles; FDG-PET. I selected these biomarkers to represent the different types of measurements that are routinely collected in AD research, and to include biomarkers that become dynamic during the pre-symptomatic and symptomatic phases of AD. MRI volumes were corrected for differences in head size by regressing against total intracranial volume (TIV). FDG-PET uptake values were averaged over the angular gyrus, inferior temporal gyrus, and posterior cingulate gyrus. For simplicity I only model baseline CSF, as modelling longitudinal CSF requires a new set of measurements for each time point to be modelled, as all the CSF measurements are re-processed once a new follow up is completed.

### 5.3.2.2 Generic Alzheimer's disease simulation model

I start from a generic model of AD, adapting the settings to perform a stability analysis of the EBM and DEM. I develop a generic model of AD that is based on the following assumptions:



- At baseline a set of time points  $\vec{t}_0$  are sampled from a uniform baseline time point distribution  $P(t_0) = \text{Unif}(0, t_r)$ , where  $t_r$  is the range of the initial subject time points.
- The follow-up time points are sampled sequentially from a set of possible follow-up times under the assumption that a proportion,  $r_f$ , of subjects drop out per year.
- The time points are assumed to follow a normal distribution centred around the time after follow-up, i.e.  $P(t|f) = \text{Norm}(f + t_0, \sigma_t)$ .
- The subtype distribution is  $P(s) = \text{Cat}(p_{s,m})$ , where  $p_s$  is the probability that a subject is assigned subtype  $s$ . By default there is only one disease subtype and so all subjects follow the typical AD set of biomarker trajectories. Alternative subtypes are used to simulate subjects that do not follow the typical AD sequence of biomarker abnormality, e.g. subjects with other neurodegenerative diseases or who are aging normally.
- The collected subset of biomarkers,  $\vec{e}$ , is sampled sequentially for the available time points by modelling an initial proportion of subjects,  $p_e$ , in which the biomarker is collected and a drop out rate per year,  $r_e$ , i.e. of the subjects that remain in the study, only a proportion of these have a measurement for a particular biomarker.
- The trajectory evolution function is sigmoidal, as has been hypothesized by Jack et al. [74], with parameters  $\vec{\theta} = (a, r, c, g)$ , where  $a$  is the trajectory minimum,  $r$  the range (difference between trajectory maximum and minimum value),  $c$  the centre point and  $g$  the gradient. To make the magnitude of the gradient a more intuitive quantity, I re-parameterise  $g$  so that it is the biomarker ‘transition time’. I define this as  $\tau = \frac{4}{g}$ , i.e. it is the time taken for the tangent to the sigmoid at the centre point,  $c$ , to transition from the minimum biomarker value,  $a$ , to the maximum biomarker value,  $a + r$ . Hence, I

have

$$z(t, \vec{\theta}) = a + \frac{r}{1 + \exp\left(-\frac{4}{\tau}(t - c)\right)},$$

with parameters  $\vec{\theta} = (a, r, c, \tau)$ .

- The trajectory parameters are normally distributed according to the trajectory parameter distributions:

- $P(a) = \text{Norm}(\mu_a, \Sigma_a)$
- $P(s) = \text{Norm}(\mu_r, \Sigma_r)$
- $P(c|s) = \text{Norm}(\mu_{c,s}, \Sigma_c)$
- $P(\tau) = \text{Norm}(\mu_\tau, \Sigma_\tau)$

The parameters  $\mu_a, \mu_r, \mu_{c,s}, \mu_\tau$  are the trajectory parameter means;  $\Sigma_a, \Sigma_r, \Sigma_{c,s}, \Sigma_\tau$  are the inter-subject covariances of the trajectory parameters.

- The measurement noise distribution follows a normal distribution with mean 0 and standard deviation  $v$ , i.e.  $P(\varepsilon) = \text{Norm}(0, v)$ .
- There are three diagnoses,  $d = \{\text{CN}, \text{MCI}, \text{AD}\}$ , as there are in ADNI, that follow a categorical distribution:  $P(d|x) = \text{Cat}(p_{\text{CN}}, p_{\text{MCI}}, p_{\text{AD}})$ , where  $p_{\text{CN}}$  is the probability that a subject is assigned a CN diagnosis,  $p_{\text{MCI}}$  is the probability that a subject is assigned a MCI diagnosis, and  $p_{\text{AD}}$  is the probability that a subject is assigned an AD diagnosis. The probability of each diagnosis  $p_d$  is evaluated using each subjects biomarker data,  $x$ , as  $p_d \propto \prod_{i=1}^I \text{Norm}(x_i, \mu_{d,i}, \sigma_{d,i})$ .

I tested the agreement between data sets generated using the default parameter values and ADNI by calculating the Bhattacharyya coefficient [169],  $BC$ , between simulated data sets and data from ADNI. The Bhattacharyya coefficient measures the similarity between two probability distributions, ranging from 0 to 1, where a Bhattacharyya coefficient of 0 corresponds to no overlap.

$$D_B = \frac{1}{8} \mu^T \Sigma^{-1} \mu + \frac{1}{2} \ln \left( \frac{\det(\Sigma)}{\sqrt{\det(\Sigma_{\text{ADNI}}) \det(\Sigma_{\text{simulated}})}} \right)$$

where

$$\mu = \mu_{\text{ADNI}} - \mu_{\text{simulated}},$$

and

$$\Sigma = \frac{\Sigma_{\text{ADNI}} + \Sigma_{\text{simulated}}}{2}.$$

The Bhattacharyya coefficient,  $BC$ , is  $BC = \exp(-D_B)$ .

I find that data sets generated using the default parameters values show good agreement with the ADNI data set, giving an average Bhattacharyya coefficient across 25 sample data sets of 0.99 when considering the biomarkers to be independent, and 0.83 when considering the dependence between biomarkers.

### 5.3.2.3 Datasets for event-based model stability analysis

I use the generic AD model to perform a set of simulations to assess how robust the EBM and DEM are to different choices of parameters.

For these experiments I assume the following set of default parameters (Table 5.1, Figure 5.2). These default parameters are intended as an idealized basis for the stability analysis, from which each parameter can be varied individually so as to explore the robustness of the models to variations in a particular parameter, independently of other effects. For each experiment I generate synthetic datasets that have 800 subjects and the biomarker set: CSF  $A\beta_{1-42}$ , CSF t-tau, CSF p-tau, FDG, MMSE, hippocampal volume, brain volume, ventricular volume.

- Baseline time points:  $t_r = 20$ , i.e. there is a range of 20 years in which a subject's baseline visit might lie.
- Follow-ups: There are 11 possible follow-up times, as there are in ADNI, at 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, and 9 years from baseline. The drop out rate per year,  $r_f$ , is estimated from ADNI as 10%.
- Time points: the standard deviation,  $\sigma_t$ , of the actual time at which each follow-up is taken is approximated from ADNI as  $\sigma_t = 0.05$  for  $t = 0.5, 1, 1.5$ , and  $\sigma_t = 0.1$  for  $t \geq 2$  (at baseline  $\sigma_t = 0$ ).
- Subtype:  $p_{s=1} = 1$ , i.e. all subjects have the same disease subtype by default.

- Biomarker collection: The proportion of subjects in which each biomarker is collected at baseline,  $p_e$ , and the drop out rate per year,  $r_e$ , are estimated from ADNI as  $p_e = 100\%, 85\%, 55\%, 50\%$ , and  $r_e = 0\%, 10\%, 5\%, 100\%$  for cognitive test scores, MRI volumes, FDG-PET hypometabolism, and CSF levels respectively (i.e. only baseline CSF modelled).
- Trajectory parameters:  $\mu_a$  is estimated from the mean biomarker value in CN subjects from ADNI (Table 5.1).  $\mu_r$  is estimated as the difference between the mean biomarker value in AD and CN subjects from ADNI. (Table 5.1) For CSF  $A\beta_{1-42}$  I use only amyloid negative CN subjects, and amyloid positive AD subjects (amyloid positive is defined as CSF  $A\beta_{1-42} < 192$  pg/ml). I chose settings for  $\mu_c$  so that the biomarkers become abnormal in the order: Abeta, P-tau, T-tau, FDG-PET, Hippocampal volume, MMSE, Ventricles, Whole brain volume (Table 5.1). I set  $\mu_\tau$  to 5 years for all biomarkers (Table 5.1).
- Trajectory inter-subject covariance:  $\Sigma_a$  is estimated from the set of 28 amyloid negative CN ADNI subjects at baseline that have measurements for all biomarkers (Table 5.1), I remove the contribution of measurement noise by subtracting the estimated measurement variance level (see next bullet point). By default I set  $\Sigma_r = 0$ ,  $\Sigma_c = 0$ ,  $\Sigma_\tau = 0$ .
- Measurement noise: I estimate the measurement noise level  $v$  for each biomarker using baseline and 6 month follow-up measurements in CN subjects from ADNI (Table 5.1) under the assumption that fluctuations in controls over a 6 month period are representative of measurement noise. For CSF I only have baseline measurements and so I set the measurement noise to 0, i.e. I model the variance as being purely inter-subject covariance rather than removing the contribution of measurement noise as I do for the other biomarkers.
- Diagnosis: I estimate  $\mu_d$  and  $\sigma_d$  for each diagnostic group using the available data for each biomarker from ADNI.

In the experiments I vary each of the following parameters of the simulations in turn, and set the rest of the parameters to their default value, generating 25 synthetic datasets for each new parameter value, and fitting the EBM to each sample dataset.

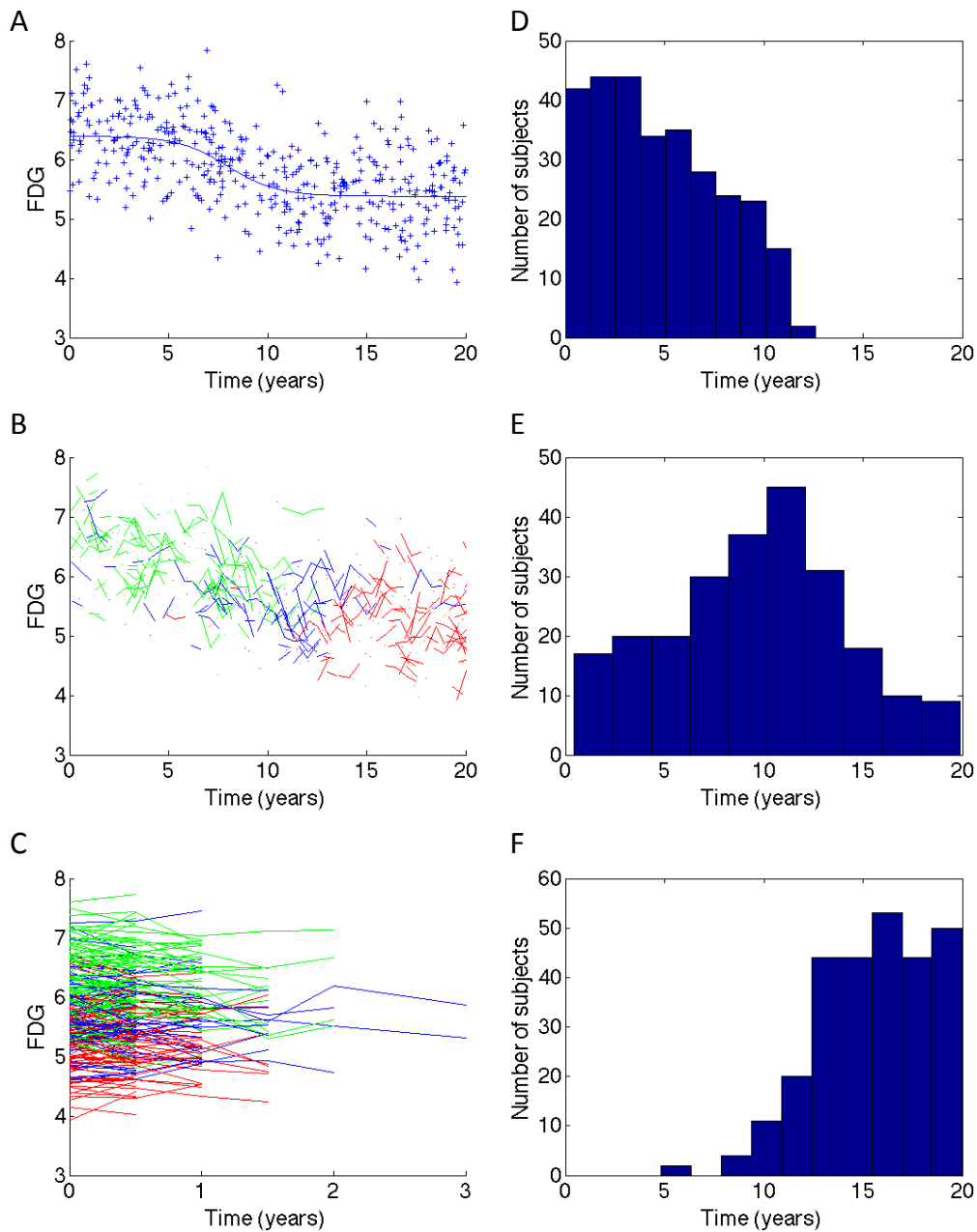
	Abeta	P-tau	T-tau	FDG	Hippo	MMSE	Vents	Brain
$\mu_a$	243	20	63	6.39	903	29.1	-7400	34300
$\mu_r$	-110	23	63	-1.01	-1630	-5.78	13600	-64000
$\mu_c$	5	6	7	8	10	12	14	15
$\mu_\tau$	5	5	5	5	5	5	5	5
$\frac{v^2}{\mu_r^2}$	0	0	0	0.030	0.006	0.013	0.004	0.022
$\frac{\Sigma_a}{\mu_r \mu_r^T}$	Abeta	0.058	0.001	-0.004	0.014	-0.001	0.002	0.013
	P-tau		0.150	0.126	0.041	-0.069	-0.015	0.048
	T-tau			0.178	0.031	-0.084	-0.003	-0.013
	FDG				0.308	0.064	0.008	-0.191
	Hippo					0.198	-0.007	-0.320
	MMSE						0.055	0.009
	Vents							1.519
	Brain							-0.896
								0.960

**Table 5.1:** Default parameter values for the EBM.  $\mu_a$  = mean trajectory minimum,  $\mu_r$  = mean trajectory range,  $\mu_c$  = mean trajectory centre point (years),  $\mu_\tau$  = mean trajectory transition time (years).  $\frac{v^2}{\mu_r^2}$  = biomarker measurement variance, normalised by the range of the trajectory.  $\frac{\Sigma_a}{\mu_r \mu_r^T}$  = covariance of the trajectory minimum, normalised by the range of the trajectory. All other covariance matrices are set to 0 by default. Reproduced from [167]. CC BY 4.0.

#### 5.3.2.4 Datasets for differential equation model stability analysis

For the DEM experiments I initially fitted a DEM to each biomarker using the default simulation settings for the EBM. However, these experiments show that the DEM does not perform well for the levels of noise estimated from ADNI. I therefore simplify the default settings for the DEM to characterise the types of noise the DEM is most sensitive to. By default I instead generate a single idealized (zero noise) biomarker trajectory with the following settings.

- Biomarker collection:  $p_e = 100\%$ ,  $r_e = 0\%$ .
- Trajectory parameters:  $\mu_a = 0$ ,  $\mu_r = 1$ ,  $\mu_c = 10$ ,  $\mu_\tau = 5$ .
- Trajectory inter-subject covariance:  $\Sigma_a = 0$ ,  $\Sigma_r = 0$ ,  $\Sigma_c = 0$ ,  $\Sigma_\tau = 0$ .



**Figure 5.2:** Example simulated FDG-PET, subject diagnosis, and time point data generated using default parameter values for the EBM (Table 5.1). (A) Simulated baseline FDG-PET data (curve shows mean trajectory); (B) simulated FDG-PET follow-up data for cognitively normal subjects (green), mild cognitive impairment subjects (blue), and Alzheimer's disease subjects (red); (C) same as (B) but plotted against follow up time rather than time point along the disease. (D)-(F) Histogram of the number of (D) cognitively normal, (E) mild cognitive impairment and (F) Alzheimer's disease subjects, at each time point at baseline. Reproduced from [167]. CC BY 4.0.

- Measurement noise:  $\nu = 0$ .

In the experiments I again vary each of the following parameters of the simulations in turn, and set the rest of the parameters to their default value, generating 25 synthetic datasets for each new parameter value, and fitting the DEM to each sample dataset.

### 5.3.3 The event-based model

The EBM [98] considers disease progression as a sequence of events at which biomarkers transition from a normal level, i.e. as seen in healthy controls, to an abnormal level, i.e. as seen in AD subjects. The maximum likelihood (ML) ordering of these events can be determined by finding the sequence  $S$  that maximizes the data likelihood

$$P(X|S) = \prod_{j=1}^J \left[ \sum_{k=0}^I \left( P(k) \prod_{i=1}^k P(x_{ij}|E_i) \prod_{i=k+1}^I P(x_{ij}|\neg E_i) \right) \right]$$

Here,  $E_i$ ,  $i = 1 \dots I$ , are events, whose occurrence is informed by the corresponding measurements  $x_{ij}$  of biomarker  $i$  in subject  $j$ ,  $j = 1 \dots J$  via the biomarker distributions: the likelihood that an event has occurred and thus the corresponding biomarker measurement  $x_{ij}$  is abnormal,  $P(x_{ij}|E_i)$ , or has yet to occur and so the corresponding biomarker measurement is normal,  $P(x_{ij}|\neg E_i)$ .  $P(k)$  is the prior likelihood of being at stage  $k$ , where events  $E_1, \dots, E_k$  have occurred, and events  $E_{k+1}, \dots, E_K$  have yet to occur. I assume no prior knowledge of disease stage by choosing the prior  $P(k)$  to be uniform. I fit a mixture of normal distributions to determine the mean,  $\mu_E$  and  $\mu_{\neg E}$ , and standard deviation,  $\sigma_E$  and  $\sigma_{\neg E}$ , of the biomarker distributions  $P(x|E)$ , and  $P(x|\neg E)$ . To guide the fitting in cases where the biomarker distributions overlap significantly, I constrain the parameters so that the standard deviation of each distribution is less than or equal to the standard deviation of biomarker measurements in the AD and control (CN) population respectively. For missing biomarker values I impute the value of  $x$  such that  $P(x|E) = P(x|\neg E)$ .

### 5.3.4 Differential equation model

I fit the DEM to each biomarker separately using a similar technique to Villemagne et al. [81]. I first calculate the rate of change in each subject by fitting a least-square linear regression to the first three available time points for each participant. For each simulation I compared fitting a linear model to the first three available time points (baseline, 0.5 years and 1 year) with fitting a linear model to the three or more available time points (up to a maximum of 9 years). I found that fitting the linear model to the first three available time points produced trajectories with the least error for all experiments, and so I only present results for fitting to the first three available time points. I fit a quadratic differential equation model (representative of the sigmoidal biomarker dynamics modelled in the simulations) to the mean biomarker value of each subject  $x_i$ , and rate of change of each subject  $\frac{\partial x_i}{\partial t}$  estimated from the linear model, i.e. I optimise for A, B and C over all subjects  $i$  such that:

$$\frac{\partial x_i}{\partial t} = Ax_i^2 + Bx_i + C$$

I then integrate this quadratic differential equation model to get the average trajectory across the population:

$$x(t) = \frac{\sqrt{4AC - B^2} \tan\left(\frac{1}{2}(k+t)\sqrt{4AC - B^2}\right) - B}{2A}$$

Where  $k$  is an unknown constant to be specified by choosing an initial condition.

### 5.3.5 Evaluation metrics

#### 5.3.5.1 Event-based model biomarker distribution parameters

Defining a ground truth for the biomarker distribution parameters when the biomarker trajectories are not binary is not straightforward, requiring the portion of the biomarker trajectory belonging to the ‘normal’ and ‘abnormal’ biomarker distribution to be defined. However, to explore the effect of the accuracy of the biomarker distribution on the estimation of the event sequence I ran each of the experiments for two settings: one where I estimated the biomarker distributions and



another where these were fixed. I fix the biomarker distributions so that  $\mu_{-E} = \mu_a$ ,  $\mu_E = \mu_a + \mu_r$ ,  $\sigma_{-E} = \sigma_E = \sqrt{\text{diag}(\Sigma_a) + v^2}$ , where  $\text{diag}(\Sigma)$  is the diagonal of the covariance matrix  $\Sigma$ .  $\mu_{a,m}$  and  $\mu_{b,m}$  are calculated from the average subject demographics. Whilst these may not be the ‘true’ biomarker distribution parameters, they provide an estimate of how the EBM behaves for a reasonable setting of the biomarker distribution parameters.

### 5.3.5.2 Event-based model Kendall’s tau distance

A key outcome measure I am interested in is the model’s ability to recover the ML sequence from the simulated data. This can be evaluated by measuring the Kendall’s tau distance (a measure of the similarity of two sequences) between the recovered sequence and a ground truth event sequence. The Kendall’s tau distance is the total number of pairwise disagreements between two sequences,  $\pi$  and  $\pi_0$ :

$$d_K(\pi, \pi_0) = \sum_{l \prec_{\pi} j} 1_{j \prec_{\pi_0} l}$$

Where  $\pi$  and  $\pi_0$  are permutations and  $l \prec_{\pi} j$  means that  $l$  precedes  $j$  in the permutation  $\pi$ . Here I use the normalized Kendall’s tau distance, i.e. I divide by the maximum distance, which is the total number of possible pairs:  $\frac{n!}{2!(n-2)!}$ , where  $n$  is the number of events in the sequence.

### 5.3.5.3 Event-based model positional variance diagrams

I use positional variance diagrams (PVDs) to look at the location of variations in the ML sequence (rather than just the extent, which I measure using the Kendall’s tau distance) for each simulation. Each entry of the positional variance diagram is the proportion of samples in which a particular event appears at that position in the ML sequence.

### 5.3.5.4 Differential equation model transition time

I compare the transition time of the simulated trajectories across the population with the ground truth simulation setting of the transition time,  $\mu_{\tau}$ .

### 5.3.6 Implementation

All experiments were performed using Matlab on a standard workstation (Intel Core i7, 3.1 GHz, 8GB memory). The event-based model event distribution parameters and maximum likelihood sequence were estimated using the same procedure as Chapters 3 and 4. All of the gaussian mixture model fits were assessed visually to ensure the solutions were physically plausible. Here to find the characteristic event sequence  $\bar{S}$ , I performed 1,000 iterations of a greedy ascent algorithm (as in [98]), initialised from 5 random start points. Less start points and iterations were sufficient as the simulated set of events is small (only eight biomarkers). I checked that the start points converged to a single maximum to ensure the global optimum had been found. The differential equation model was fitted using the Matlab ‘polyfit’ function, the integrated model fits were computed analytically.

## 5.4 Results

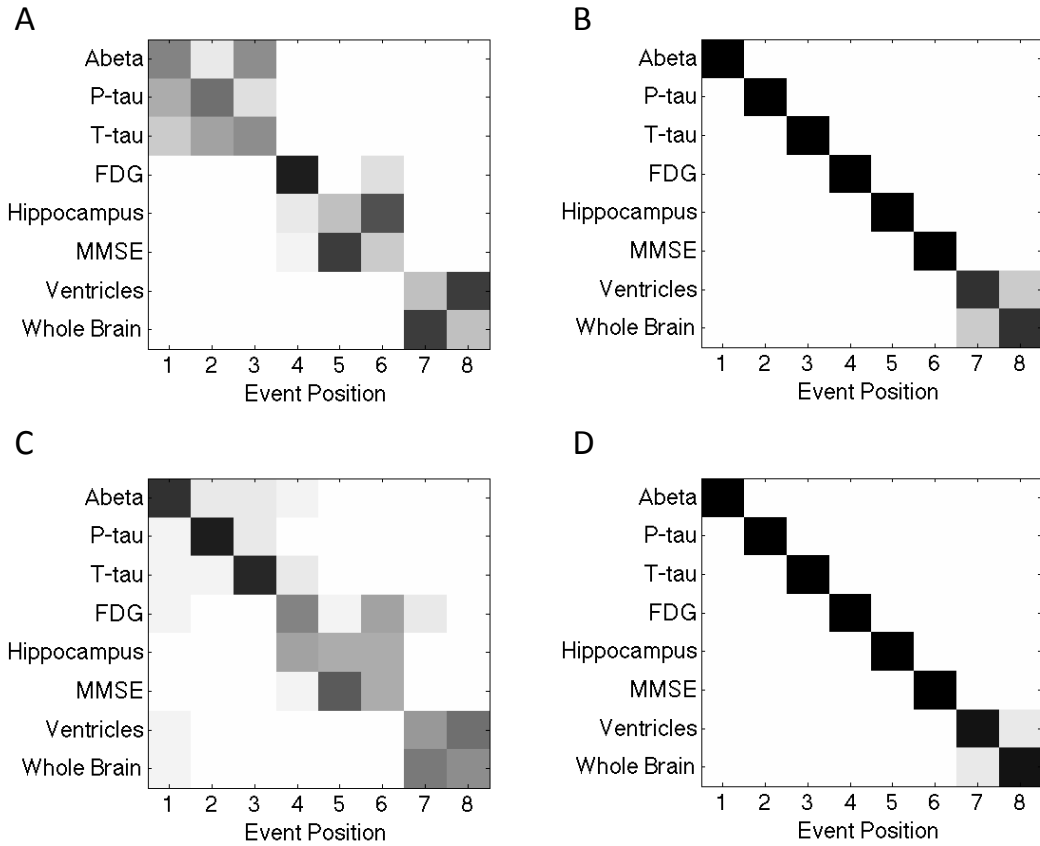
### 5.4.1 Event-based model stability analysis

I performed a stability analysis of the EBM to test how robust the model is to different types of heterogeneity that are likely to exist in sporadic AD datasets. As outlined in Section 5.3.5.1, I consider two scenarios: (1) estimated event distributions and (2) fixed event distributions. Scenario (1) is the typical scenario for fitting the event-based model to sporadic Alzheimer’s disease. In this scenario both the distribution of normal and abnormal biomarker values (the event distributions) and the sequence of biomarker abnormality events need to be estimated. Scenario (2) is a more idealistic scenario in which I fix the event distributions to a reasonable value using the parameters of the simulated data. Together these two scenarios allow the effect of the accuracy of the event distributions on the accuracy of the sequence to be disentangled.

#### 5.4.1.1 Default parameter values

Fitting the EBM to datasets generated using the default parameter values (Figure 5.3) gives a Kendall’s tau distance of  $0.11 \pm 0.05$  when estimating the event distributions from the data, and  $0.01 \pm 0.01$  for fixed event distributions. Repeating

this experiment without any missing biomarker values only slightly improves the Kendall's tau distance despite the increase in the number of data points;  $0.10 \pm 0.09$  for estimated event distributions, and  $0.00 \pm 0.01$  for fixed event distributions, showing that imputing the data such that  $P(x|E) = P(x|\neg E)$  (see Section 5.3.3) is a valid technique for fitting the EBM to data with missing biomarker values. It is worth noting that in all the simulations the inter-subject variation setting, which is estimated from ADNI, may be larger than the actual level of inter-subject variation. This is because CN subjects in ADNI may originate from a range of underlying time points along the biomarker trajectories.



**Figure 5.3:** Results of applying the EBM to synthetic data with missing values (A)-(B) and without missing values (C)-(D) generated using the default parameters. In (A) and (C) the event distributions are estimated, and in (B) and (D) the event distributions are fixed. Reproduced from [167]. CC BY 4.0.

### 5.4.1.2 Experiment 1: Noise levels

I ran five simulations to look at the effect of different levels of: (A) measurement noise  $v$ , (B) inter-subject covariance of the trajectory minimum  $\Sigma_a$ , (C) inter-subject covariance of the trajectory range  $\Sigma_r$ , (D) inter-subject covariance of the trajectory centre  $\Sigma_{c,s}$ , (E) inter-subject covariance of the trajectory transition time  $\Sigma_\tau$ . For each of the respective simulations I vary the noise level as a proportion  $p$  of (A) the estimated measurement noise from ADNI (see Section 5.3.2.3 and Table 5.1), and as a proportion  $p^2$  of the covariance matrices: (B) the estimated inter-subject covariance of the trajectory minimum from ADNI (see Section 5.3.2.3 and Table 5.1), (C) the square of the mean of the trajectory range  $\mu_r^2$  estimated from ADNI (see Section 5.3.2.3 and Table 5.1), (D) the range of the trajectory centre points squared,  $10^2$  years, (E) the range of the baseline time points squared,  $20^2$  years. For simulations (C), (D), and (E) I assume a diagonal covariance matrix. Varying the measurement noise  $v$  (Table 5.2 A: Experiment A) has little effect on the Kendall's tau distance between the sample event sequences and the ground truth as the estimated measurement noise level is small compared to the inter-subject covariance of the trajectory minimum (Table 5.1). Varying the inter-subject covariance of the trajectory minimum  $\Sigma_a$  (Table 5.2 A: Experiment B) has a large effect on the Kendall's tau distance for the estimated event distributions, but little effect for fixed event distributions. This shows that it is difficult to estimate the parameters of the event distributions for high biomarker inter-subject variance levels. For very low variance levels on  $\Sigma_a$  the Kendall's tau distance increases again for both fixed and estimated event distributions. This is probably because at very low variance the event distributions don't model the biomarker values over the central portion of the biomarker trajectory, where the trajectory transitions from the minimum to the maximum value. This makes it ambiguous as to whether the biomarker is normal or abnormal during the trajectory transition, making it difficult for the EBM to order the biomarkers. The EBM is robust to inter-subject variation in the trajectory range  $\Sigma_r$  (Table 5.2 A: Experiment C), giving a similar Kendall's tau distance to the default settings for noise levels up to 50% of the range  $\mu_r$ . The EBM is quite

robust to variation in the trajectory centre points  $\Sigma_{c,s}$  (Table 5.2 A: Experiment D) and transition time  $\Sigma_{\tau}$  (Table 5.2 A: Experiment E).

#### A. Noise Levels

		0.01	0.1	0.5	1	2
Estimated	A	0.10(0.06)	0.09(0.05)	0.11(0.06)	<b>0.10(0.05)</b>	0.12(0.05)
	B	0.12(0.06)	0.00(0.01)	0.02(0.02)	<b>0.09(0.05)</b>	0.36(0.16)
	C	<b>0.11(0.05)</b>	0.11(0.06)	0.13(0.06)	0.20(0.17)	0.18(0.17)
	D	<b>0.09(0.05)</b>	0.13(0.04)	0.24(0.17)	0.32(0.12)	0.47(0.18)
	E	<b>0.10(0.05)</b>	0.09(0.04)	0.15(0.10)	0.26(0.12)	0.40(0.11)
Fixed	A	0.00(0.01)	0.00(0.01)	0.01(0.02)	<b>0.01(0.01)</b>	0.01(0.02)
	B	0.17(0.04)	0(0)	0.00(0.01)	<b>0.00(0.01)</b>	0.07(0.04)
	C	<b>0.00(0.01)</b>	0.01(0.01)	0.03(0.04)	0.11(0.04)	0.17(0.10)
	D	<b>0.00(0.01)</b>	0.01(0.01)	0.05(0.04)	0.09(0.05)	0.17(0.05)
	E	<b>0.01(0.02)</b>	0.01(0.01)	0.01(0.02)	0.10(0.06)	0.19(0.06)

#### B. Trajectory Parameters

		1	5	10	15	20
Estimated	A	0.44(0.21)	0.19(0.04)	<b>0.09(0.04)</b>	0.08(0.15)	0.20(0.22)
	B	0.03(0.02)	<b>0.11(0.05)</b>	0.19(0.04)	0.24(0.05)	0.30(0.12)
Fixed	A	0.38(0.22)	0.09(0.04)	<b>0.01(0.01)</b>	0(0)	0.01(0.01)
	B	0.00(0.01)	<b>0.00(0.01)</b>	0.01(0.03)	0.04(0.04)	0.05(0.03)

#### C. Time Sampling

		0	1	2.5	5
Estimated		0.12(0.06)	0.10(0.06)	0.11(0.07)	<b>0.10(0.04)</b>
Fixed		0(0)	0.00(0.01)	0(0)	<b>0.00(0.01)</b>

#### D. Subtypes

		0	0.25	0.5	0.75	1
Estimated		<b>0.10(0.04)</b>	0.17(0.20)	0.50(0.36)	0.91(0.06)	0.96(0.05)
Fixed		<b>0.01(0.01)</b>	0.04(0.04)	0.45(0.39)	0.95(0.04)	1.00(0.01)

**Table 5.2:** Mean (standard deviation of the mean in brackets) Kendall’s tau distance between ground truth event sequence and event sequence returned by the EBM for synthetic data with varying: (A) noise levels, (B) trajectory parameters, (C) time sampling, (D) subtypes. In all tables each column is a new parameter value, and ‘estimated’ and ‘fixed’ refer to whether the event distributions are estimated by the EBM or fixed to known values. Bold values represent the default parameters of the EBM. In part (A), experiments A-E represent varying: A. measurement noise; B. inter-subject covariance of trajectory minimum; C. inter-subject covariance of trajectory range; D. inter-subject covariance of trajectory centre; E. inter-subject covariance of trajectory transition time. In part (B), experiments A-B represent varying: A. trajectory centre points, B. trajectory transition times. Reproduced from [167]. CC BY 4.0.

### 5.4.1.3 Experiment 2: Trajectory parameters

I performed two experiments to test the robustness of the EBM to different values of the trajectory parameters: (A) varying the centre points of the trajectories,  $\mu_c$ , and (B) varying the transition time of the trajectories  $\mu_\tau$ . In (A) I assume an evenly spaced set of trajectory centre points over a segment of the disease time course. I vary the duration of this segment as a fraction of 10 years. The trajectory centre points are centred about the middle point along the disease time course (10 years). In (B) I vary the transition time as a fraction of the overall range of the disease time course (20 years), keeping the transition time the same for all biomarkers. The EBM has difficulty estimating the event sequence for both fixed and estimated event distributions when the trajectory centres are close together (Table 5.2 B: Experiment A), and for longer transition times (Table 5.2 B: Experiment B), which violate the assumption of the EBM that an event has either occurred or not occurred. For estimated event distributions the EBM also has difficulty ordering the events when the trajectory centres are spread over the full disease time course (Table 5.2 B: Experiment A). This is because the portion of the trajectory where the biomarker is normal (for early biomarkers) or abnormal (for late biomarkers) is not observed. As in the previous experiments, fixing the event distributions improves the estimation of the event sequence.

### 5.4.1.4 Experiment 3: Time sampling

In this experiment I look at how under-sampling of the disease time course affects the ability of the EBM to recover the sequence of biomarker abnormality. I assume that the time points are sampled from a mixture of three Gaussian distributions, with means at 5 years, 10 years and 15 years respectively. I vary the standard deviation of these distributions, assuming that all of the gaussians have the same standard deviation. This allows simulation of CN, MCI and AD subjects being at entirely different points along the disease time course. The EBM is robust to under-sampling of the disease time course (Table 5.2 C), giving a similar Kendall's tau distance to the default settings for all simulations.

#### 5.4.1.5 Experiment 4: Subtypes

To explore the effect of including a set of subjects that follow a different event sequence I modelled two disease subtypes, varying the fraction of subjects that belong to each subtype. For both subtypes the trajectory centre points are evenly spaced from a minimum of 5 years to a maximum of 15 years. In subtype 1 the biomarkers become abnormal in the same order as the default settings: Abeta, P-tau, T-tau, FDG-PET, hippocampal volume, MMSE, ventricles, whole brain volume. In subtype 2 the biomarkers become abnormal in the reverse sequence. This sequence has a Kendall's tau distance of 1 from the sequence of subtype 1. The EBM is robust up to a proportion of 25% of subjects that follow an alternative event sequence (Table 5.2 D): at 25% outliers the Kendall's tau distance is similar to the result for 0% outliers for estimated event distributions, and only slightly increased for fixed event distributions. Likewise, at 75% outliers, when the majority of subjects are subtype 2 the Kendall's tau distance is only slightly worse than for 100% outliers. At 50% outliers the EBM alternates between estimating a sequence similar to subtype 1 and subtype 2.

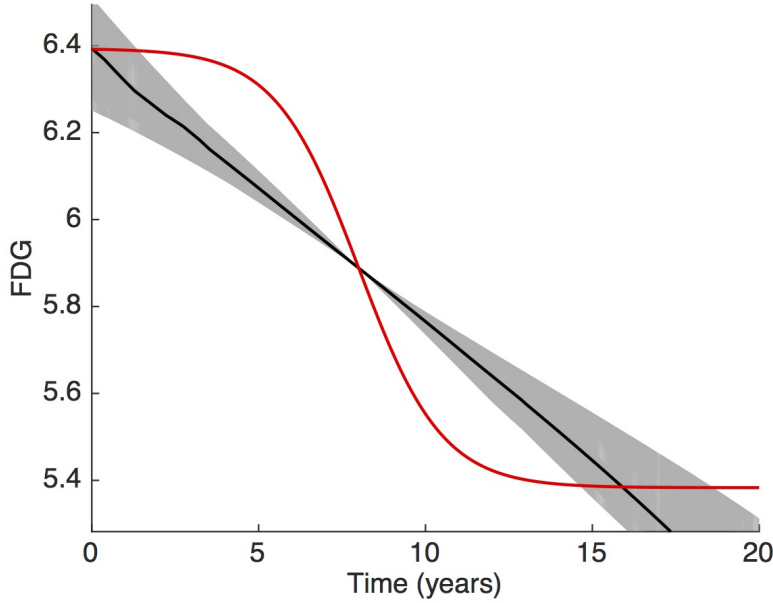
### 5.4.2 Differential equation model stability analysis

I performed a stability analysis of the DEM to test how robust the model is to varying the noise levels on the trajectory parameters.

#### 5.4.2.1 Default parameter values

Fitting a DEM to each biomarker in turn (see Figure 5.4 for estimated synthetic trajectories for FDG-PET) using the default parameter values for the EBM gives an average across biomarkers (excluding CSF for which I only modelled baseline collection) of a mean sample transition time  $12.8 \pm 1.1$  years and standard deviation of this sample transition time of  $2.2 \pm 1.2$  years. This is more than double the simulated trajectory transition time of 5 years. For biomarkers with more data points available (higher biomarker collection rate, e.g. MMSE) the standard deviation of the transition time reduces but the mean transition time remains similar, i.e. the DEM becomes more confident in the biased estimate of the trajectory transition

time. In subsequent experiments I simplify the default parameter settings for the DEM to a zero noise case (see Section 5.3.2.4) to characterise the types of parameter noise that the DEM is most sensitive to.



**Figure 5.4:** Integrated DEM trajectories for FDG-PET uptake generated using synthetic data with the default settings for the EBM (Table 5.1). The ground truth trajectory is in red, and the median estimated trajectory is in black with the inter-quartile range shaded in grey. Reproduced from [167]. CC BY 4.0.

#### 5.4.2.2 Stability of the differential equation model to noise

For the DEM default parameter values (zero noise case) the DEM is able to recover the trajectory transition time much more accurately (sample transition time is  $5.1 \pm 0.0$  years for a simulated trajectory with a transition time of 5 years). As with the EBM, I ran five simulations to look at the effect of different levels of: (A) measurement noise  $v$ , (B) inter-subject variance of the trajectory minimum  $\Sigma_a$ , (C) inter-subject variance of the trajectory range  $\Sigma_r$ , (D) inter-subject variance of the trajectory centre  $\Sigma_{c,s}$ , (E) inter-subject variance of the trajectory transition time  $\Sigma_\tau$ . I vary the noise level as a proportion  $p$  of (A)  $\mu_r$ , and a proportion  $p^2$  of the variance: (B)  $\mu_r^2$ , (C)  $\mu_r^2$ , (D)  $\mu_\tau^2$ , (E)  $\mu_\tau^2$ . The DEM is sensitive to measurement noise (Table 5.3: Experiment A), and variance of the trajectory minimum (Table 5.3: Experiment B) and range (Table 5.3: Experiment C), with 25% measurement noise giving a



transition time of around twice as long as the actual transition time. The DEM is less sensitive to variance of the trajectory transition time (Table 5.3: Experiment E). The DEM is unaffected by noise in the trajectory centre point (Table 5.3: Experiment D), as this is removed by differentiating.

	0.1	0.25	0.5	1
A	5.9(0.4)	9.8(3.1)	12.7(4.7)	15.0(5.2)
B	7.5(0.2)	11.4(0.5)	13.3(0.5)	14.4(0.7)
C	6.6(0.2)	9.8(0.5)	11.6(0.7)	12.3(1.0)
D	5.1(0.0)	5.1(0.0)	5.1(0.0)	5.1(0.0)
E	5.1(0.0)	5.1(0.1)	5.4(0.3)	8.7(0.9)

**Table 5.3:** Mean (standard deviation of the mean in brackets) estimated transition time (years) for DEM fitted to synthetic data with varying: A. measurement noise; B. inter-subject variance of trajectory minimum; C. inter-subject variance of trajectory range; D. inter-subject variance of trajectory centre; E. inter-subject variance of trajectory transition time. For all simulations the ground truth transition time is 5 years. Each column is a new parameter value. Reproduced from [167]. CC BY 4.0.

## 5.5 Discussion

I have presented a framework for the simulation of sporadic neurodegenerative disease datasets. I applied the framework to generate synthetic AD data, and thereby provide insight into the robustness of the EBM and a DEM to the likely variation in sporadic disease datasets.

### 5.5.1 Simulation framework

The simulation framework I have presented is simple and flexible. For example, it is easily extendible to include subjects with a range of demographics, for example age, gender, and education, or genetic risk factors. Such effects can be modelled as a transformation of the trajectory parameters. Here I simplify the diagnosis as a relationship with the biomarker values, however, a more realistic diagnosis procedure could be simulated that is based on, for example, cognitive test results. It is also possible to add in a screening procedure that post-selects subjects with a similar set of demographics to the dataset being simulated, for example age matching across diagnostic categories. Here I only consider inter-subject variance, however,

intra-subject variance for longitudinal datasets could also be modelled.

## 5.5.2 Stability analysis

### 5.5.2.1 Event-based model stability analysis

The EBM stability analysis shows that the EBM is sensitive to the estimation of the event distribution parameters. However, when the event distribution parameters are estimated accurately, the EBM is very robust to the likely heterogeneity in sporadic disease datasets. I find that the EBM is robust to noise in the trajectory parameters, different choices of trajectory parameters, under-sampling of the underlying disease time course, and outliers who follow different event sequences.

### 5.5.2.2 Differential equation model stability analysis

For all simulations the DEM under estimates the trajectory gradient leading to an over estimation of the trajectory transition time. For the level of noise estimated from the ADNI data this over estimate is more than twice as long as the ground truth trajectory transition time. Whilst this result may be in part due to an over estimation of the amount of inter-subject variation from ADNI, the stability analysis of the DEM shows the DEM will severely over-estimate the trajectory transition time even when the inter-subject variation is much lower. The DEM is very sensitive to measurement noise and inter-subject variation of the trajectory minimum (normal biomarker level) and range (difference between a normal and abnormal level). I further find that using three time points to fit the DEM rather than all available time points, for which the approximation to the derivative is less valid, gives a better estimate of the trajectory transition time, even under high noise levels.

### 5.5.2.3 Limitations

In the set of experiments presented I vary each parameter in turn. However, there will likely be multiplicative effects of varying these parameters in combination. I further make a set of assumptions that are specific to hypothetical models of AD, such as sigmoidal trajectories, and to the design of the ADNI dataset, such as the proportion of subjects that drop out per year, and the proportion of subjects in which each biomarker is collected. I also assume that measurement errors are Gaussian,

which may not be the best choice of noise distribution for all of the biomarkers. Therefore, although these simulations do provide an insight into the types of effects which can be expected from different datasets, the simulations should be re-run with dataset specific parameters to assess the performance of the EBM and DEM on alternative datasets. I chose the DEM to be similar to [81], however other DEM approaches (e.g. [131]) may recover a more accurate estimate of the trajectory transition time for heterogeneous data sets, and should be tested in future work.

### 5.5.3 Implications for the application and development of data-driven models

#### 5.5.3.1 The event-based model

The results of the EBM stability analysis show that the EBM is sensitive to the accuracy of the estimated biomarker distribution models,  $P(x|E)$ , and  $P(x|\neg E)$ . Therefore the application of the EBM is most effective when the biomarkers have distinct control and case distributions. The results further show that the EBM is robust to 25% outlier corruption, which is higher than the proportion of misdiagnoses I expect in typical sporadic neurodegenerative disease cohorts. The simulations highlight several key areas for improvement of the EBM. First, better estimation techniques for the biomarker distribution parameters should aid recovery of the event sequence when the control and case distributions are not well defined. Adaptation of the EBM to take into account the uncertainty in the biomarker distribution parameters, e.g. by sampling the distribution parameters simultaneously with the ordering, may also help to ameliorate this problem. Second, although the EBM can estimate the ML event sequence for a modest proportion of outliers, it is unable to distinguish other likely event sequences in the data. Future work will look at fitting mixture models with multiple event sequence modes to the data.

#### 5.5.3.2 Differential equation models

The simulations show that the DEM is sensitive to noise, leading to over estimation of the trajectory transition time, meaning that the DEM should only be applied to biomarkers with low measurement noise and inter-subject variance. Alternatively,

robust fitting techniques need to be developed that can correct for the bias encountered when fitting a DEM to noisy biomarker trajectories. The simulations further show that it is important that the duration of follow up for each individual is a good approximation to the derivative (short with respect to the full disease time course). This is shown to be more important than the inclusion of lots of follow up time points, which improves the accuracy of the estimated derivative, suggesting that follow up data over a longer time period should be discarded when fitting a DEM.

## 5.6 Conclusion

I have presented a framework for generating synthetic neurodegenerative disease datasets (contribution 1.3.2 (a), Chapter 1, page 40), which can be used to evaluate the robustness of data-driven models to likely variations in sporadic disease datasets, and to directly compare them. I have demonstrated the use of this framework to evaluate the stability of the EBM and a DEM of disease progression to heterogeneity in the ADNI dataset (contribution 1.3.2 (b), Chapter 1, page 40). Future work will use the simulation framework to evaluate the stability of other data-driven models, such as self-modelling regression approaches [128]. The simulation framework can further be used as a technique for validating extensions to data-driven models, to determine model weaknesses, and to highlight areas for improvement.

## Chapter 6

# Multiple orderings of events in disease progression

This chapter presents an initial formulation of the dynamic clustering technique developed in Chapter 7. The work presented in this chapter was presented at the *Information Processing in Medical Imaging* conference in 2015, the proceedings of which were published in *Lecture Notes in Computer Science*. I developed the methodology, performed the analysis and wrote the manuscript; my co-authors provided feedback on the methodology and the manuscript. My co-author Razvan Marinescu made the visualisations for Figures 6.2, 6.3 and 6.5.

## 6.1 Associated publications

A. L. Young, N. P. Oxtoby, J. Huang, R. V. Marinescu, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander. Multiple Orderings of Events in Disease Progression. In *Information Processing in Medical Imaging*, volume 9123 of *Lecture Notes in Computer Science*, pages 711–722, 2015

## 6.2 Introduction

The sequence in which biomarkers become abnormal provides a simple, intuitive description of disease progression, giving insights into the underlying disease biology and a potential mechanism for disease staging. The sequence of biomarker abnormality in sporadic neurodegenerative diseases, e.g. AD, has been a topic of

intense debate amongst neurologists [74]. Reconstructing this sequence for sporadic neurodegenerative diseases is difficult because the position of subjects with respect to the full disease time course is unknown. Typically clinical diagnoses are used as a time proxy, but this limits the temporal resolution of the sequence, e.g. in AD there are usually only three clinical diagnosis categories: cognitively normal (CN), mild cognitive impairment (MCI) and AD [85]. Additional complications arise due to the long disease time course [81] and inherent heterogeneity of sporadic disease datasets. Many different factors contribute to this heterogeneity [75, 146], for example genetic disease subtypes, mixed pathology, environmental factors, and misdiagnosed subjects.

The event-based model (EBM) [98] considers disease progression as a series of events, where each event corresponds to a new biomarker becoming abnormal. By considering cross-sectional patient data as snapshots of a single common event sequence, the EBM is able to probabilistically reconstruct the ordering of events across subjects, without relying on a-priori disease staging. Taking samples of the posterior probability of this sequence provides insight into the uncertainty in this single event ordering. The application of this model has been demonstrated in dominantly-inherited AD and Huntington's disease [98] to determine the sequence in which regional brain volumes become abnormal, and in sporadic AD to determine the sequence in which cerebrospinal fluid (CSF) markers, cognitive test scores, and a limited set of regional atrophy and brain volume biomarkers become abnormal [134]. Young et al. [134] found that this sequence is different in APOE4 positive individuals, with increased genetic risk of AD, compared to the whole population, suggesting that the whole population contains a proportion of subjects who do not follow the single ordering of events encoded by the EBM.

The assumption made by the EBM in [98] and [134] of a single ordering of events in all subjects is a major simplification for heterogeneous sporadic disease datasets. In this work I relax this assumption by considering a family of models that allow for multiple and distributed orderings of events. The first is a generalised Mallows model [171], which parameterises the variance in the single order-

ing, allowing subjects to deviate from the central event sequence. The second is a Dirichlet process mixture model [172], which allows for subgroups of subjects that follow different event sequences. Previous work [157] on generalised Mallows EBMs relied on a well-defined control population and a complete set of biomarkers for each subject. Here I re-formulate this model to remove the reliance on a well-defined control population, allowing the model to be fitted to heterogeneous sporadic disease datasets, and to handle missing data, providing a multi-modal picture of disease progression. I formulate a Gibbs sampling technique that further provides samples of the uncertainty in the model parameters. I additionally introduce a new model: Dirichlet process mixtures of generalised Mallows EBMs, and develop a Gibbs sampler to estimate its parameters [173]. I apply these models to determine the sequence in which FDG-PET, CSF markers, cognitive test scores, and a large set of regional brain volumes become abnormal in sporadic AD.

## 6.3 Models

### 6.3.1 The event-based model

The EBM of disease progression [98] consists of a set of events  $\{e_1, \dots, e_N\}$  and an ordering  $\sigma = (\sigma(1), \dots, \sigma(N))$ , where  $\sigma(k) = i$  means that event  $e_i$  occurs in position  $k$ . In practise I only observe a snapshot of the event sequence for each subject, taken at an unknown stage  $k$ . If a subject is at stage  $k$  in the sequence  $\sigma$  the events  $e_{\sigma(1)} \dots e_{\sigma(k)}$  have occurred and events  $e_{\sigma(k+1)} \dots e_{\sigma(N)}$  have yet to occur. This adduces a partition of the event set, or partial ranking,  $\gamma_k = e_{\sigma(1)}, \dots, e_{\sigma(k)} | e_{\sigma(k+1)}, \dots, e_{\sigma(N)}$ , where the vertical bar indicates that the first set of events precedes the second. The occurrence of event  $e_i$  in subject  $j$  is informed by biomarker measurement  $x_{ij}$ . The generative model of the biomarker data is

$$k_j \sim P(k),$$

$$x_{\sigma(i),j} \sim p(x_{\sigma(i),j} | e_{\sigma(i)}) \text{ if } i \leq k_j,$$

$$x_{\sigma(i),j} \sim p(x_{\sigma(i),j} | \neg e_{\sigma(i)}) \text{ otherwise.}$$

$p(x|e)$  and  $p(x|\neg e)$  are probability density functions on observing biomarker measurement  $x$  given that event  $e$  has or has not occurred respectively.  $P(k)$  is a prior on the disease stage  $k$ .

### 6.3.2 The generalised Mallows event-based model

I formulate the generalised Mallows EBM by using a generalised Mallows model to parameterise the variance in a central event sequence  $\pi$  through the spread parameter  $\vec{\theta} = (\theta_1, \dots, \theta_{N-1})$ . Each subject then has their own latent ordering  $\sigma_j$ , which is assumed to be a sample from a generalised Mallows model. The generative model of the biomarker data in the EBM is therefore preceded by

$$\pi, \vec{\theta} \sim P(\pi, \vec{\theta} | \mathbf{v}, \vec{r}),$$

$$\sigma_j \sim GM(\pi, \vec{\theta}).$$

$GM(\pi, \vec{\theta}) = \frac{1}{\psi(\vec{\theta})} \exp[-d_{\vec{\theta}}(\pi, \sigma)]$  is a generalised Mallows distribution with  $\psi(\vec{\theta}) = \prod_{j=1}^{n-1} \psi_{n-j}(\theta_j) = \prod_{j=1}^{n-1} \frac{1-e^{-(n-j+1)\theta_j}}{1-e^{-\theta_j}}$ .  $d_{\vec{\theta}}(\pi, \sigma)$  is the generalised Kendall's tau distance [171], which penalises the number of pairwise disagreements between sequences.  $P(\pi, \vec{\theta} | \mathbf{v}, \vec{r})$  is a conjugate prior over the generalised Mallows distribution parameters of the form  $P(\pi, \vec{\theta} | \mathbf{v}, \vec{r}) \propto \exp(-\mathbf{v} \sum_j [\theta_j r_j + \ln \psi_{n-j}(\theta_j)])$  [174].

### 6.3.3 Dirichlet process mixtures of generalised Mallows event-based models

Dirichlet process mixtures of generalised Mallows models assume that each subject has their own central ordering  $\pi_j$  and spread parameters  $\vec{\theta}_j$ , which are sampled from a discrete distribution  $G$  that is drawn from a Dirichlet process [172]. A Dirichlet process mixture is a generative clustering model where the number of clusters is a random variable, meaning that the number of clusters is detected automatically depending on the concentration parameter  $\alpha$ . The generative model of the biomarker data in the EBM is now preceded by the process

$$G \sim DP(\alpha, P(\pi, \vec{\theta} | \mathbf{v}, \vec{r})),$$



$$\pi_j, \vec{\theta}_j \sim G,$$

$$\sigma_j \sim GM(\pi_j, \vec{\theta}_j),$$

where  $DP(\alpha, P(\pi, \vec{\theta} | \mathbf{v}, \vec{r}))$  is a Dirichlet process [172]. Each data point  $\pi_j$  can be characterised by an association with a cluster label  $c_j \in 1, \dots, C$  and each cluster  $c$  with a set of generalised Mallows parameters  $\sigma_c$  and  $\vec{\theta}_c$ .

## 6.4 Inference

### 6.4.1 The event-based model

Inference in the EBM can be performed by taking Markov Chain Monte Carlo (MCMC) samples of  $P(\sigma|X) = \frac{P(X|\sigma)P(\sigma)}{P(X)}$  where

$$P(X|\sigma) = \prod_{j=1}^J \left[ \sum_{k=0}^K P(k) \left( \prod_{i=1}^k p(x_{\sigma(i),j} | e_{\sigma(i)}) \prod_{i=k+1}^N p(x_{\sigma(i),j} | \neg e_{\sigma(i)}) \right) \right]. \quad (6.1)$$

### 6.4.2 The generalised Mallows event-based model

I use Gibbs sampling to infer the parameters of the generalised Mallows EBM. This consists of two stages. First, generating a set of sample event sequences  $\sigma_{1:J}$ . I sample from an augmented model [157], by alternating between sampling a subject's ordering  $\sigma_j$  and disease stage  $k_j$ , which are used to deterministically reconstruct their partial ranking  $\gamma_j$ . The Gibbs sampling updates are therefore

$$\sigma^{(j)} \sim P(\sigma | \vec{\gamma} = \gamma_j, \pi, \vec{\theta}),$$

$$k^{(j)} \sim P(k | \vec{\sigma} = \sigma_j, X_j).$$

Second, sampling the model parameters given the set of sample orderings  $\sigma_{1:J}$  using the updates

$$\pi \sim P(\pi | \vec{\theta}, \mathbf{v}, \vec{r}, \sigma_{1:J}),$$

$$\theta_k \sim P(\theta_k | \pi, \mathbf{v}, \vec{r}, \sigma_{1:J}).$$

### 6.4.3 Dirichlet process mixtures of generalised Mallows event-based models

I formulate another Gibbs sampler to infer the parameters of Dirichlet process mixtures of generalised Mallows EBMs. I generate a set of candidate sample orderings  $\sigma_{1:J,1:C}$ , disease stages  $k_{1:J,1:C}$ , and partial rankings  $\gamma_{1:J,1:C}$ , which are conditioned on the parameters for each cluster via the updates

$$\sigma^{(j,c)} \sim P(\sigma | \vec{\gamma} = \gamma_{jc}, \pi_c, \vec{\theta}_c),$$

$$k^{(j,c)} \sim P(k | \vec{\sigma} = \sigma_{jc}, X_j).$$

From these samples I sample the cluster assignment  $c_j$  of each subject conditioned on the cluster assignments of the other subjects  $c_{-j}$ , where  $c_{-j}$  is the set of cluster assignments for all subjects except subject  $j$ , the subject's sample ordering for each cluster  $\sigma_{j,1:C}$ , disease stage  $k_{j,1:C}$  and their biomarker data  $X_j$ . I then update the generalised Mallows model parameters for each cluster,  $\pi_c$  and  $\vec{\theta}_c$ , from the set of subject orderings assigned to each cluster,  $\vec{\sigma}_c$ . So I have the updates

$$c^{(j)} \sim P(c | c_{-j}, \sigma_{j,1:C}, \vec{\theta}, \alpha, \nu, \vec{r}, X_j, k_{j,1:C}),$$

$$\pi^{(c)} \sim P(\pi | \vec{\theta} = \vec{\theta}_c, \nu, \vec{r}, \vec{\sigma}_c),$$

$$\theta_k^{(c)} \sim P(\theta_k | \vec{\pi} = \pi_c, \nu, \vec{r}, \vec{\sigma}_c).$$

## 6.5 Implementation

### 6.5.1 ADNI dataset

I considered 382 subjects (135 CN subjects, 149 MCI, 98 AD) who had a 1.5T structural MRI (T1) scan at baseline. I calculated the total volume (left plus right hemisphere) of 82 regions in the Neuromorphometrics parcellation<sup>1</sup> for each subject, correcting for head size variance by regressing against total intracranial volume. Seg-

---

<sup>1</sup><http://neuromorphometrics.org:8080/>

mentation was performed using the Geodesic Information Flow framework [175]. I retained the 35 regions having significant differences between CN and AD subjects using the Wilcoxon rank sum test with  $p < 0.01$ . I downloaded biomarker values from the ADNI database (adni.loni.usc.edu) for CSF markers ( $A\beta_{1-42}$ , tau, p-tau), cognitive test scores (MMSE, RAVLT, ADAS-Cog), and global FDG-PET metabolism.

### 6.5.2 Model fitting

I compare the result of fitting the EBM, generalised Mallows EBM and Dirichlet process mixtures of generalised Mallows EBMs to the ADNI data set for the set of 42 biomarker abnormality events described. Following previous work [98] I model the probability that a biomarker is normal,  $p(x|\neg e)$ , as a Gaussian distribution, and the probability that a biomarker is abnormal,  $p(x|e)$ , as a uniform distribution covering the full range of observed values to reflect the range of severity that corresponds to an abnormal biomarker. I use a mixture model to fit these distributions to the data to account for a proportion of outliers in the control population, and visually assess the fit of the Gaussian mixture models. In subjects that had missing data points I imputed the biomarker values such that  $p(x|e) = p(x|\neg e)$ , i.e. it is equally probable that the event  $e$  has or has not occurred. The prior probability that a subject is at a particular disease stage  $P(k)$  is assumed to be uniform. To fit the generalised Mallows model I need to sample  $\sigma$  from  $P(\sigma|\gamma, \pi, \vec{\theta})$ . I approximate this by sampling from a generalised Mallows model for each of the event sets in the partial ranking  $\gamma$  separately; the set of events  $\gamma_e$  that have occurred and the set of events  $\gamma_{\neg e}$  that have yet to occur. I sample  $\sigma_e \sim GM(\pi_{\gamma_e}, \vec{\theta}_{\gamma_e})$ , and  $\sigma_{\neg e} \sim GM(\pi_{\gamma_{\neg e}}, \vec{\theta}_{\gamma_{\neg e}})$ . This means that the precedence of events specified by the partial ranking is preserved, and that the central ordering of the generalised Mallows model for each event set,  $\pi_{\gamma_e}$  and  $\pi_{\gamma_{\neg e}}$ , has the minimal Kendalls tau distance [171] from the central ordering  $\pi$  of the full generalised Mallows model. I sample  $k$  from  $P(k|\sigma, X_j)$  using equation 6.1, i.e.  $P(k|\sigma, X_j) \propto \prod_{i=1}^k p(x_{\sigma(i),j}|e_{\sigma(i)}) \prod_{i=k+1}^N p(x_{\sigma(i),j}|\neg e_{\sigma(i)})$ . The remaining sampling updates follow the algorithm in [173]. I sample  $\pi$  exactly using a stage-wise algorithm, and  $\vec{\theta}$  using a beta function approximation. I used the

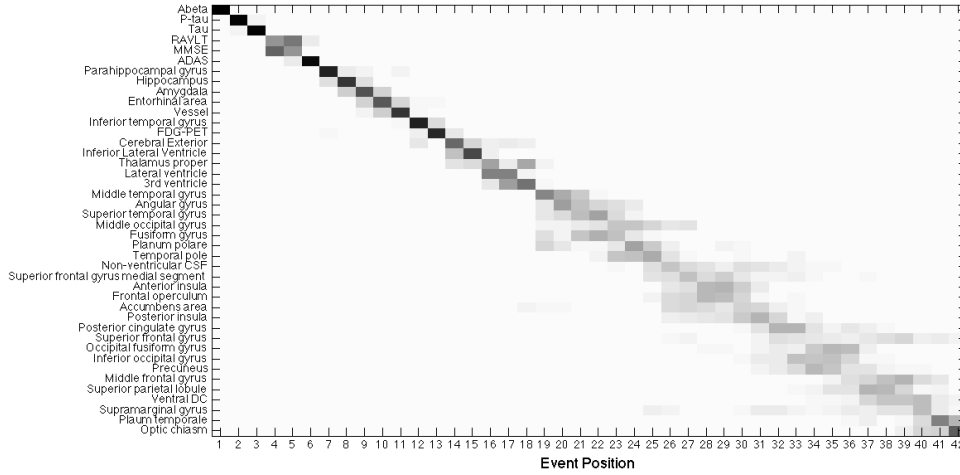
Beta-Gibbs algorithm [173] to update the Dirichlet process mixture model cluster assignments  $c_j$ , weighting the probability each subject belongs to each cluster by  $P(X_j|\sigma_{j,c},k_{j,c})$ , and the generalised Mallows model parameters  $\pi_c, \vec{\theta}_c$  for each cluster. I fix the priors to be  $\nu = 1, \vec{r} = \vec{1}, \alpha = 1$ . I initialise  $\pi$  randomly,  $\gamma_e$  as the set of events with  $p(x|e) > p(x|\neg e)$ ,  $\gamma_{\neg e}$  as the set of events with  $p(x|e) \leq p(x|\neg e)$ , and the Dirichlet process mixture to have 25 clusters. The computational time required for the generalised Mallows EBM and the Dirichlet process mixture of generalised Mallows EBM grows rapidly as the number of subjects and biomarkers increases because of the requirement to update several individual subject parameters at each iteration, in addition to estimating overall population parameters for each cluster. For the present experiments each iteration takes around a minute using Matlab on a standard workstation (Intel Core i7, 3.1 GHz, 8GB memory).

## 6.6 Results and discussion

### 6.6.1 The event-based model

Figure 6.1 shows a positional variance diagram of the MCMC samples of the single ordering of events returned by the EBM. I visualise a few key stages of this sequence in the top row of Figure 6.3 to show the spatial correspondence of the sequence of regional volume loss estimated by the model. I find that CSF markers are the first to become abnormal, followed by cognitive test scores, then memory-related brain regions, then FDG-PET, and then other AD-related brain regions. This sequence complements the findings of other studies, but provides a much more detailed picture of the regional progression of volume changes than has been seen previously in sporadic AD, and a direct comparison of the sequence of regional changes relative to a multi-modal set of biomarkers. Fonteijn et al. [98] looked at the regional progression of volume loss but in familial AD and using atrophy rates. The results in Young et al. [134] show a multi-modal sequence of biomarker abnormality in sporadic disease but for a small set of regional volumes, and hippocampal and whole brain atrophy rates from short-term longitudinal MRI. Here I show the first multi-modal sequence of biomarker abnormality in sporadic AD, including a large set of

regional volumes. I am able to construct this picture from entirely cross-sectional data, and incorporate biomarkers with missing values.

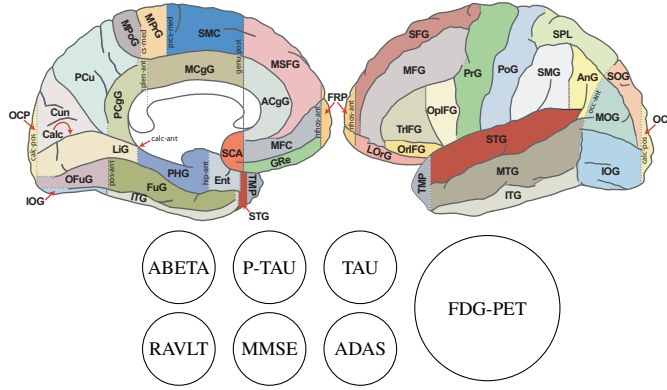


**Figure 6.1:** Central ordering estimated by the EBM: Positional variance diagram of the MCMC samples of the maximum likelihood event sequence  $\sigma$ . The events on the y-axis are ordered by the maximum likelihood sequence estimated by the model. Each entry of the positional variance diagram represents the proportion of samples in which a particular event appears in a particular position in the central ordering, ranging from 0 in white to 1 in black. A black diagonal corresponds to high certainty in the ordering of events, whereas grey blocks in the diagram mean that the events permute. Reproduced from [170] with permission from Springer. The final publication is available at Springer via [http://dx.doi.org/10.1007/978-3-319-19992-4\\_56](http://dx.doi.org/10.1007/978-3-319-19992-4_56).

### 6.6.2 The generalised Mallows event-based model

The generalised Mallows EBM estimates both the central ordering of the events and the variance in this single event ordering across the population (Figure 6.3). Figure 6.3 compares the central ordering  $\pi$  and variance  $\vec{\theta}$  estimated by the generalised Mallows EBM, i.e. the range of event sequences across the population, with the central ordering estimated by the EBM. The central event sequence has a similar ordering to the EBM, but the variance in this central ordering of events increases, as shown by the increase in the number of orange regions in Figure 6.3. By using Gibbs sampling I further obtain estimates of the uncertainty in each of the model parameters, as well as the latent variables included in the model, for example a subject's disease stage (Figure 6.4). Fitting the generalised Mallows EBM means that

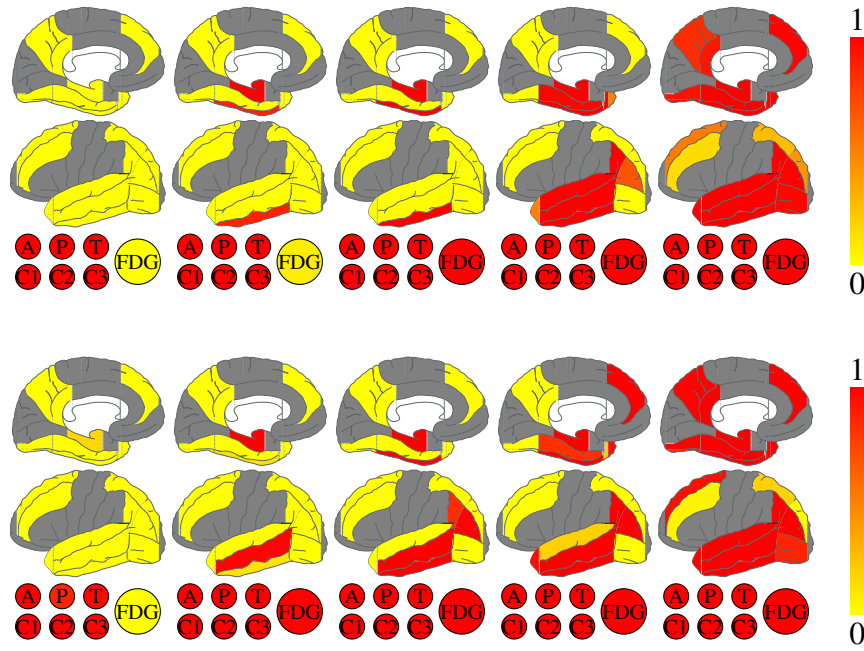
the uncertainty in this stage accounts for the variance in the ordering of the events across the population.



**Figure 6.2:** Key for Figures 6.3 and 6.5, generated using the BrainColorMap software<sup>2</sup>. Reproduced from [170] with permission from Springer. The final publication is available at Springer via [http://dx.doi.org/10.1007/978-3-319-19992-4\\_56](http://dx.doi.org/10.1007/978-3-319-19992-4_56).

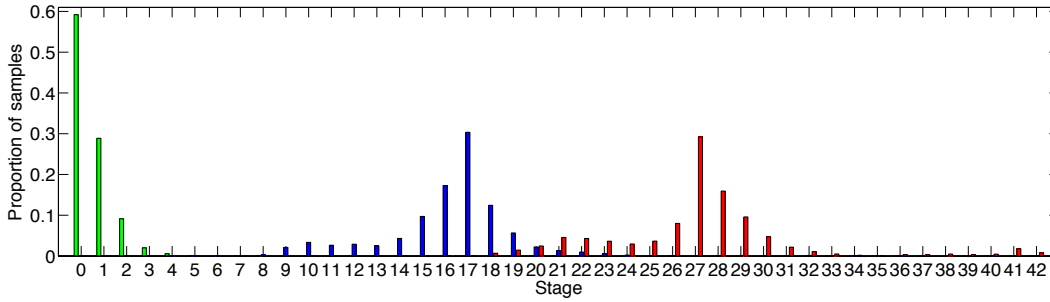
### 6.6.3 Dirichlet process mixtures of generalised Mallows event-based models

I fitted a Dirichlet process mixture of generalised Mallows EBM to allow for clusters of subjects that follow different sequences of events, of which each cluster has its own central ordering  $\pi_c$  and variance  $\vec{\theta}_c$ . The Dirichlet process mixture model identifies three main clusters in the data, with an average proportion of  $0.48 (\pm 0.02)$ ,  $0.24 (\pm 0.10)$ , and  $0.29 (\pm 0.10)$  subjects being assigned to each cluster respectively over the Gibbs samples. Figure 6.5 compares the estimated central ordering and variance for each of the clusters. The first two clusters look more AD-like than the third cluster, producing a similar event sequence to the EBM and generalised Mallows model (Figure 6.3), with CSF biomarkers and memory-related brain regions becoming abnormal early in the sequence. The third cluster likely captures outliers that do not fit the AD sequence of events. The ordering of events for the third cluster consists of only mild cognitive deficits and no CSF abnormalities, perhaps representing a normal aging event sequence, or simply reflecting that regional volume loss is a noisy measure on a cross-sectional level. The variance  $\vec{\theta}_c$  is greater for the clusters of the Dirichlet process mixture model than the gen-



**Figure 6.3:** Comparison of the central ordering estimated by the event-based model (top) with the generalised Mallows model (bottom) (see key in Figure 6.2). We display the results for six stages: stage 6, 12, 18, 24 and 36, where each stage number corresponds to the number of biomarkers that have become abnormal. Each biomarker (brain region, CSF, cognitive test or FDG-PET) is coloured according to the proportion of the population in which it has become abnormal by a particular stage along the central ordering. This proportion is estimated for the event-based model by the number of MCMC samples (Figure 6.1), and for the generalised Mallows model by the probability (calculated using the central ordering  $\pi$  and spread  $\vec{\theta}$ ) of an event appearing at or before a particular stage. This proportion ranges from 0 in yellow to 1 in red. Regions not included in the model are shown in grey. At each stage yellow biomarkers can be interpreted as being normal, red biomarkers as being abnormal, and orange biomarkers as varying in whether they have become abnormal across the population. Reproduced from [170] with permission from Springer. The final publication is available at Springer via [http://dx.doi.org/10.1007/978-3-319-19992-4\\_56](http://dx.doi.org/10.1007/978-3-319-19992-4_56).

eralised Mallows model (as shown by an increase in the number of orange regions in Figure 6.5 compared to Figure 6.3), likely because each cluster only contains a proportion of the population, meaning that there are fewer subjects to fit the model to, and due to the uncertainty in the cluster assignment of each subject. The Gibbs sampling technique returns samples of all of the model parameters. For example, I am able to estimate the uncertainty in the disease stage of each subject for both models, and the cluster assignment of each subject from the Dirichlet process mixture,



**Figure 6.4:** Estimate of the uncertainty in a subject’s disease stage obtained by using Gibbs sampling to fit the generalised Mallows EBM. I show an estimate of the probability of each stage for an example CN subject (green), MCI subject (blue), and AD subject (red). Each stage corresponds to the number of biomarkers in the sequence that have become abnormal. Reproduced from [170] with permission from Springer. The final publication is available at Springer via [http://dx.doi.org/10.1007/978-3-319-19992-4\\_56](http://dx.doi.org/10.1007/978-3-319-19992-4_56).

producing a similar diagram to Figure 6.4.

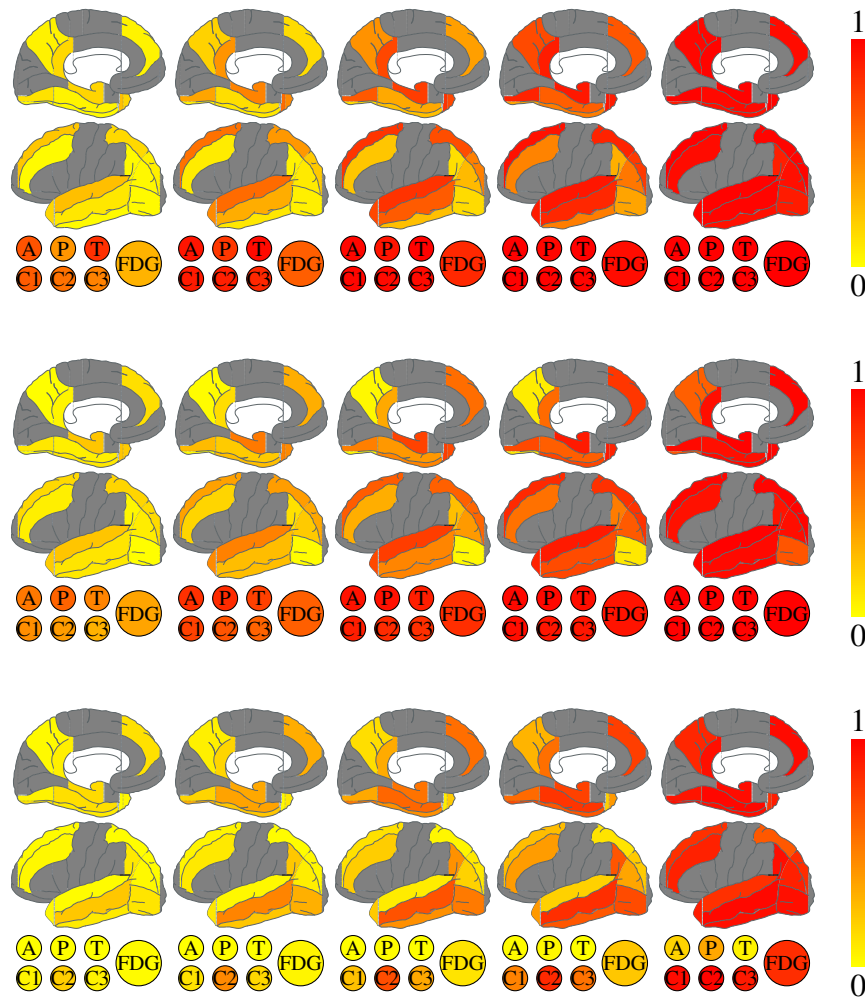
## 6.7 Conclusion

I proposed a generalised family of EBMs that relax the assumption of a common event sequence over the population in different ways. I formulated these models so that they work for a large multi-modal set of sporadic disease biomarkers, and developed a Gibbs sampler that provides an estimate of the uncertainty on each model parameter. I fitted this family of models to the ADNI dataset to determine the ordering of a much more extensive, multi-modal set of biomarkers than has been seen previously. I find that the generalised Mallows model estimates a similar event sequence to the original EBM, but with a larger variation across subjects. Fitting a Dirichlet process mixture model detects subgroups of the population with different event sequences.

The models developed in this chapter provide an interesting first attempt at modelling heterogeneity in biomarker progression patterns. However, a major concern is that the variance parameter  $\vec{\theta}$  of the generalised Mallows EBMs (and the variance parameter  $\vec{\theta}_c$  of the Dirichlet process mixture of generalised Mallows EBMs), greatly increases model complexity with no real gain in clinical utility. The parameter  $\vec{\theta}$  describes the variance in the progression pattern across the population,

<sup>2</sup><http://braincolor.mindboggle.info/>.





**Figure 6.5:** As Figure 6.3, but for the clusters identified by the Dirichlet process mixture of generalised Mallows event-based models (top to bottom: clusters 1 to 3). Reproduced from [170] with permission from Springer. The final publication is available at Springer via [http://dx.doi.org/10.1007/978-3-319-19992-4\\_56](http://dx.doi.org/10.1007/978-3-319-19992-4_56).

which will always be very large given that subjects are only observed at a single time point, and therefore the possible sequences they could have gone through to reach that point are all of the possible permutations of the events that have occurred at that time point. Estimating this variance doesn't provide any additional information that is clinically useful.

In the subsequent chapter I develop a more parsimonious model of heterogeneous biomarker progression patterns.



## **Chapter 7**

# **A data-driven model of disease subtypes with distinct patterns of biomarker evolution in frontotemporal dementia and Alzheimer's disease**

This chapter details contribution 1.3.3 (see Chapter 1, page 41). My colleague Razvan Marinescu made the visualisations for Figures 7.1, 7.2, 7.3, 7.4 and 7.5, and Supplementary Figures A.5 and A.12.

### **7.1 Introduction**

Neurodegenerative diseases, such as AD and frontotemporal lobar degeneration (FTLD), are often heterogeneous in their pathology, genetics and clinical presentation. The identification of data-driven disease sub-populations and quantification of their patterns of biomarker evolution can elucidate relationships between different biomarker measurements that contribute to this observed heterogeneity, providing insights into underlying disease mechanisms. Moreover, quantitative models of the evolution of biomarkers in different disease subtypes can be used for precision medicine by providing a quantitative tool for patient stratification and prognostica-

tion, with application in clinical trials and healthcare.

Frontotemporal dementia (FTD) is the clinical presentation of FTLN, a highly heritable set of neurodegenerative diseases that are characterised by progressive frontal and temporal lobe atrophy. FTD comprises three major clinical subtypes: behavioural variant FTD, semantic dementia, and primary progressive aphasia. Around a third of FTD cases are inherited on an autosomal dominant pair, with mutations in progranulin (GRN), microtubule-associated protein tau (MAPT) and chromosome 9 open reading frame 72 (C9orf72) being the most common causes. FTLN is pathologically heterogeneous, consisting of neuronal inclusions containing abnormal forms of either tau, TDP-43, or fused in sarcoma (FUS) proteins, with each of these proteinopathies having further sub classifications. Of the major familial subgroups, patients with GRN mutations have TDP-43 type A pathology, patients with MAPT mutations having tauopathies, and patients with expansions in C9orf72 have type A or type B TDP-43 pathology [176].

AD is characterised pathologically by the accumulation of amyloid plaques and neurofibrillary tangles in brain tissue. These pathologies are thought to trigger subsequent neuropathological processes leading to downstream neurodegeneration, affecting the medial temporal lobe structures first, with the progression of atrophy reflecting the pattern of increase in neurofibrillary tangles. This neurodegeneration gives rise to the clinical symptoms experienced by a patient, the earliest of which is typically memory loss. However, the pattern of neurofibrillary tangles observed at autopsy is heterogeneous, with 25% of patients not having the typical distribution of neurofibrillary tangles. This observation has led to the suggestion that there are three pathological subtypes of AD: typical, hippocampal-sparing, and limbic-predominant [27]. In hippocampal-sparing AD neurofibrillary tangles are relatively numerous in the cortex and reduced in the hippocampus, whereas at the other end of the spectrum, in limbic-predominant AD, neurofibrillary tangle counts are greater in the hippocampus and reduced in the cortex.

Pathological subtypes at post-mortem can be associated with distinct ante-mortem biomarker profiles, such as characteristic patterns of brain volume loss on

magnetic resonance imaging (MRI) [154]. In FTLD, each of the proteinopathies has shown a fairly distinct atrophy pattern on MRI [177, 178]. In genetic FTD, distinct patterns of atrophy are observed in the different mutation types [179], in addition to pathology type [178]. In AD, more severe cortical atrophy has been observed in patients with hippocampal-sparing AD, compared to more severe medial temporal lobe atrophy in limbic-predominant AD [154].

Reconstructing patterns of biomarker evolution from *in vivo* data is challenging in the presence of heterogeneity in both the disease process and the disease stage. The majority of previous studies reconstructing longitudinal patterns of biomarker evolution assume a single common progression pattern across subjects, and so cannot characterise heterogeneous progression patterns. Estimating longitudinal patterns of biomarker evolution in neurodegenerative diseases is in itself challenging, as the duration of individual follow-up is short-term, typically a few years, in comparison to the long time course the diseases emerge over, which is estimated to be over a decade. One approach to estimating longitudinal biomarker patterns from cross-sectional data is to regress longitudinal biomarker values against clinical diagnosis [104–106], or some other clinical staging measure [49, 80, 112], such as a cognitive test score. The temporal resolution of such approaches is limited by the accuracy of the staging measure, typically to only a few stages. Data-driven models of disease progression (e.g. [98, 127, 128, 132, 134]) have been proposed to allow the estimation of fine-grained biomarker evolution patterns by realigning subject’s biomarker measurements to a data-driven time axis. However, all of these studies assume that subjects belong to a single biomarker evolution pattern.

Previous attempts to disentangle heterogeneity in the disease process of AD or FTD [180–186] focus on clustering data from a single static disease stage. Such studies are fundamentally limited by the quality of the staging measure used to select subjects, and can only provide insights into disease subtypes at one stage in the disease process. Additionally, this stage may not be common across all the subtypes as the choice of disease staging measure may be more sensitive to one subtype than another.

The work of Guerrero et al. [112] reconstructs longitudinal disease progression patterns that incorporate heterogeneity in biomarker trajectories, however the time axis of their model is not data-driven, instead being indexed by time to clinical diagnosis. This means that the temporal resolution of their model is limited by the accuracy of the diagnosis, and that the model can only be applied to subjects who convert between diagnostic categories, and so requires a large amount of follow-up data, reducing subject numbers. Additionally their model does not directly propose a categorisation of the data into subgroups. In Young et al. [170], a data-driven subtyping model was proposed that describes disease progression as groups of subjects that have distinct orderings of biomarker abnormality events. However, these biomarker abnormality events constitute a binary switch, from a normal to an abnormal level, meaning that more complex dynamic behaviour of different biomarkers cannot be estimated.

In this work I present a dynamic clustering model that integrates clustering and data-driven disease progress estimation to characterise neurodegenerative disease subtypes with distinct patterns of biomarker evolution. I first validate the technique by demonstrating the ability of the model to recover known genetic subtypes of FTD using volumetric MRI data from the Genetic Frontotemporal dementia Initiative (GENFI). I show that the dynamic clustering model is able to recover the cascade of volumetric MRI loss for GRN, MAPT and C9orf72 mutation carriers without knowledge of the genetic labels. In addition, the model provides new insights into the disease process in the C9orf72 mutation carriers by revealing that this group is best modelled as two subgroups with distinct patterns of volume loss. I further apply the dynamic clustering technique to AD using volumetric MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to determine data-driven subtypes of AD and their patterns of volume loss. I demonstrate the ability of the model to provide fine-grained staging and subtyping information that independently contributes to predicting conversion from mild cognitive impairment (MCI) to AD.

## 7.2 Methods

### 7.2.1 Data description

#### 7.2.1.1 GENFI dataset

I used cross-sectional volumetric MRI data from GENFI<sup>1</sup> to fit the dynamic clustering model. Subjects were included from the second data freeze of GENFI which in total consisted of 365 participants recruited across 13 centres in the United Kingdom, Canada, Italy, Netherlands, Sweden, and Portugal. 324 had a usable volumetric T1-weighted MRI scan for analysis (15 participants did not have a scan and the other participants were excluded as the scans were of unsuitable quality due to motion, other imaging artefacts, or pathology unlikely to be attributed to FTD). The 324 participants included 144 non-carriers, 129 unaffected carriers, and 51 affected carriers. Of the 129 unaffected presymptomatic mutation carriers there were 64 GRN, 41 C9orf72, and 24 MAPT carriers. Of the 51 affected symptomatic carriers, there were 14 GRN, 26 C9orf72, and 11 MAPT carriers. The acquisition and post-processing procedures for GENFI have been previously described in [187]. Briefly, cortical and subcortical volumes were generated using a multiatlas segmentation propagation approach [175], combining cortical regions of interest to calculate grey matter volumes of the entire cortex, separated into the frontal, temporal, parietal, occipital, cingulate, and insula cortices. Because the dynamic clustering model expresses disease progression patterns in terms of z-scores relative to a control population, there is no requirement to pre-select the biomarkers according to whether they have disease signal. I therefore included all of the regions of interest in the study.

#### 7.2.1.2 ADNI dataset

I downloaded data from LONI<sup>2</sup> on 11 May 2016 and constructed two volumetric MRI datasets: those with higher (3T) and lower (1.5T) field strength. The inclusion criteria for the 3T and 1.5T datasets were having cross-sectional FreeSurfer volumes available that passed overall quality control from either a 3T (processed

---

<sup>1</sup>[www.genfi.org.uk](http://www.genfi.org.uk)

<sup>2</sup>[www.loni.ucla.edu/ADNI/](http://www.loni.ucla.edu/ADNI/)

using FreeSurfer Version 5.1) or a 1.5T (processed using FreeSurfer Version 4.3) MRI scan. The particular regional volumes selected for the study were chosen to be similar to those in GENFI. The 3T dataset consisted of 793 subjects (183 cognitively normal (CN), 86 significant memory concern (SMC), 243 early mild cognitive impairment (EMCI), 164 late mild cognitive impairment (LMCI), 117 AD), of which 73 were enrolled in ADNI-1, 99 were enrolled in ADNI-GO, and 621 were enrolled in ADNI-2. The 1.5T dataset consisted of 576 ADNI-1 subjects (180 CN, 274 LMCI, 122 AD). I corrected the volumes for variations in head size by regressing against total intracranial volume (TIV). I further downloaded age, sex, education, and APOE genotype from the ADNImerge table. I downloaded follow-up information to test the association of the dynamic clustering model subtypes and stages with longitudinal outcomes, consisting of diagnostic follow-up data and cognitive test scores from the mini-mental state examination (MMSE). I also downloaded baseline CSF measurements of  $A\beta_{1-42}$ , which I used to identify a control population.

### 7.2.1.3 Z-scores

I expressed each regional volume measurement as a z-score relative to a control population: in GENFI I used data from all non-carriers, in ADNI I used amyloid-negative CN subjects, defined as those with a CSF  $A\beta_{1-42}$  measurement greater than 192 pg/mL [139]. This gave a control population of 48 amyloid-negative CN subjects for the 3T dataset, and 56 amyloid-negative CN subjects for the 1.5T dataset. I used these control populations to determine whether the effects of age, sex or education were significant, and if so to regress them out. I then normalised each dataset relative to its control population, so that the control population had a mean of 0 and standard deviation of 1. Because regional brain volumes decrease over time the z-scores become negative with disease progression, so for simplicity I took the negative value of the z-scores so that the z-scores would increase as the brain volumes became more abnormal.



## 7.2.2 Mathematical modelling overview

### 7.2.2.1 Mathematical model for dynamic clustering

I formulate a probabilistic generative model of disease progression that consists of groups of subjects with distinct patterns of biomarker evolution. The biomarker evolution of each subgroup is described as a series of events, where each event corresponds to a biomarker reaching a particular z-score compared to a control group. This model is based on the event-based model (EBM) in [98, 134, 170], but reformulates the events so that they represent the continuous linear accumulation of a biomarker from one z-score to another, rather than an instantaneous switch from a normal to an abnormal level. The resulting model describes the biomarker evolution of each subgroup as a piecewise linear trajectory, with a constant noise level that is derived from a control population. For details of the mathematical formulation of the model and model fitting procedure see Mathematical model and Model fitting. The model assumes a fixed number of clusters  $C$ , for which I estimate the proportion of subjects  $f$  that belong to each cluster, and the order  $S_c$  in which biomarkers reach each z-score for each cluster  $c = 1 \dots C$ . I determine the optimal number of clusters  $C$  for a particular dataset through 10-fold cross-validation (see Cross-validation).

### 7.2.2.2 Uncertainty estimation

In addition to estimating the most probable sequence  $S_c$  for each cluster, I can determine the relative likelihood of all sequences for each cluster by evaluating the probability of each possible sequence. This provides an estimate of the uncertainty in the ordering  $S_c$ , which I summarise by plotting positional variance diagrams of the probability that each z-score event appears at each position in the sequence for each cluster. In practise the number of sequences is too large to evaluate all possible sequences so I use Markov Chain Monte Carlo (MCMC) sampling to provide an approximation to this uncertainty, as in [98, 134].

### 7.2.2.3 Cross-validation

I performed 10-fold cross validation of the dynamic clustering results by dividing the data into 10 folds and re-fitting the model to each subset of the data, with one

of the folds retained for testing each time. I evaluated the optimal number of clusters using the Cross-Validation Information Criterion [188], i.e. by evaluating the likelihood of each  $c$ -cluster model from  $c = 1 \dots C$  on the test data for each fold and choosing the model with the highest out-of-sample likelihood across all folds.

#### 7.2.2.4 Patient subtyping and staging

I assigned subjects to subtypes and stages predicted by the dynamic clustering model by evaluating the likelihood that they belonged to each stage of each subtype and choosing the combination of subtype and stage with the highest likelihood. When evaluating the likelihood I integrated over the set of MCMC samples to account for the uncertainty in the model parameters, rather than just evaluating the likelihood at the maximum likelihood parameters. This means that a patient's model stage indicates the average number of z-score events in the sequence that have occurred. In the experiments I added in a normal-appearing cluster to separate out subjects with normal-appearing brain volumes, i.e. to avoid the situation where subjects with normal appearing brain volumes have equal probability of belonging to the initial stage of any subtype. I classified subjects with a stage of 0 as normal-appearing, but harsher model stage thresholds can be used for a more conservative classification.

#### 7.2.2.5 Static clustering

I compared the dynamic clustering model to a static clustering model, which consists of a mixture of gaussians with unknown mean and variance using expectation-maximisation. The dynamic clustering model depends on a known control population so to ensure a fair comparison between static and dynamic clustering I included a fixed normal-appearing cluster in the static clustering model, with a mean of 0 and a standard deviation of 1 for all biomarkers. As for the dynamic clustering model, I evaluated the optimal number of clusters using the Cross-Validation Information Criterion [188].

### 7.2.3 Mathematical model

The dynamic clustering model is formulated as a mixture of linear z-score EBMs. The linear z-score EBM is based on a continuous generalisation of the EBM in [98, 134], which I describe first.

The EBM in [98, 134] describes disease progression as a series of events, where each event corresponds to a biomarker transitioning from a normal to an abnormal level. The occurrence of an event,  $E_i$ , for biomarker  $i = 1 \dots I$ , is informed by the measurements  $x_{ij}$  of biomarker  $i$  in subject  $j$ ,  $j = 1 \dots J$ . The whole dataset  $X = \{x_{ij} | i = 1 \dots I, j = 1 \dots J\}$  is the set of measurements of each biomarker in each subject. The most likely ordering of the events is the sequence  $S$  that maximises the data likelihood

$$P(X|S) = \prod_{j=1}^J \left[ \sum_{k=0}^I \left( P(k) \prod_{i=1}^k P(x_{ij}|E_i) \prod_{i=k+1}^I P(x_{ij}|\neg E_i) \right) \right], \quad (7.1)$$

where  $P(x|E_i)$  and  $P(x|\neg E_i)$  are the likelihoods of measurement  $x$  given that biomarker  $i$  has or has not become abnormal, respectively.  $P(k)$  is the prior likelihood of being at stage  $k$ , at which the events  $E_1, \dots, E_k$  have occurred, and the events  $E_{k+1}, \dots, E_I$  have yet to occur. The prior  $P(k)$  is assumed to be uniform. The likelihoods  $P(x|E_i)$  and  $P(x|\neg E_i)$  are modelled as normal distributions.

In this work I reformulate the EBM in (7.1) by replacing the instantaneous normal to abnormal events with events that represent the linear accumulation of a biomarker from one z-score to another. The linear z-score EBM consists of a set of  $N$  z-score events  $E_{iz}$ , which indicate the linear accumulation of biomarker  $i = 1 \dots I$  to a z-score  $z_{ir} = z_{i1} \dots z_{iR_i}$ , i.e. each biomarker is associated with its own set of z-scores, and so  $N = \sum_i R_i$ . Each biomarker is additionally associated with a maximum z-score,  $z_{\max}$ , which it accumulates to at the end of stage  $N$ . I consider a continuous time axis,  $t$ , which I choose to go from  $t = 0$  to  $t = 1$  for simplicity (the scaling is arbitrary). At each disease stage  $k$ , which has a duration from  $t = \frac{k}{N+1}$  to  $t = \frac{k+1}{N+1}$ , a z-score event  $E_{iz}$  occurs. The biomarkers evolve as time  $t$  progresses

according to a piecewise linear function  $f_i(t)$ , where

$$f(t) = \begin{cases} \frac{z_1}{t_{E_{z_1}}} t & \text{for } 0 < t \leq t_{E_{z_1}} \\ z_1 + \frac{z_2 - z_1}{t_{E_{z_2}} - t_{E_{z_1}}} (t - t_{E_{z_1}}) & \text{for } t_{E_{z_1}} < t \leq t_{E_{z_2}} \\ \vdots & \\ z_{R-1} + \frac{z_R - z_{R-1}}{t_{E_{z_R}} - t_{E_{z_{R-1}}}} (t - t_{E_{z_{R-1}}}) & \text{for } t_{E_{z_{R-1}}} < t \leq t_{E_{z_R}} \\ z_R + \frac{z_{\max} - z_R}{1 - t_{E_{z_R}}} (t - t_{E_{z_R}}) & \text{for } t_{E_{z_R}} < t \leq 1 \end{cases}$$

To clarify, the times  $t_{E_{iz}}$  are determined by the position of the z-score event  $E_{iz}$  in the sequence  $S$ , so if event  $E_{iz}$  occurs in position  $k$  in the sequence then  $t_{E_{iz}} = \frac{k+1}{N+1}$ .

To formulate the model likelihood for the linear z-score EBM I replace (7.1) with

$$P(X|S) = \prod_{j=1}^J \left[ \sum_{k=0}^N \left( \int_{\frac{k}{N+1}}^{\frac{k+1}{N+1}} \left( P(t) \prod_{i=1}^I P(x_{ij}|t) \right) dt \right) \right], \quad (7.2)$$

where,

$$P(x_{ij}|t) = \text{NormPDF}(x_{ij}, f_i(t), \sigma_i).$$

$\text{NormPDF}(x, \mu, \sigma)$  is the normal probability distribution function, with mean  $\mu$  and standard deviation  $\sigma$ , evaluated at  $x$ . I assume the prior on the disease stage  $P(t)$  is uniform, i.e.  $P(t) = 1$ , as in the original EBM.

The overall model,  $M$ , is a mixture of linear z-score event-based models, hence we have

$$P(X|M) = \sum_{c=1}^C f_c P(X|S_c)$$

here,  $C$  is the number of clusters (subtypes), and  $f$  is the proportion of subjects assigned to a particular cluster (subtype)

## 7.2.4 Model fitting

### 7.2.4.1 Hierarchical estimation

I fit the dynamic clustering model hierarchically by initialising the fitting of the  $C$  cluster model from the  $C - 1$  cluster model, i.e. I solve the clustering problem

sequentially from  $C = 1 \dots C_{\max}$ , where  $C_{\max}$  is the maximum number of clusters I would like to fit, initialising each model using the previous model. To fit the  $C$  cluster model using the  $C - 1$  cluster model, I generate  $C - 1$  candidate  $C$  cluster models by going through each of the  $C - 1$  clusters in turn and finding their optimal split into two clusters, I then use this two cluster solution together with the other  $C - 2$  clusters to initialise the fitting of the  $C$  cluster model. To optimise the  $C$  cluster model I perform an expectation maximisation algorithm, alternating between updating the sequences  $S_c$  for each cluster and the fraction  $f_c$ . Of these  $C - 1$  candidate  $C$  cluster models, I choose the model with the highest likelihood as the solution to the clustering problem.

#### 7.2.4.2 Two-cluster estimation

To find the optimal split of a cluster into two clusters, I initialise the assignments of data points to the two clusters randomly, find the optimal model parameters for these two data subsets, and use these cluster parameters to initialise the fitting of the two clusters. I repeat this procedure for different random cluster assignments until the algorithm converges to the maximum likelihood solution.

#### 7.2.4.3 Single cluster estimation

To find the optimal model parameters (the sequence  $S$  in which the biomarkers reach each z-score) for a single cluster I perform a greedy procedure whereby I initialise the sequence  $S$  randomly and then I go through each z-score event  $e$  in turn and find its optimal position in the sequence relative to the other z-score events, i.e. I fix the order of the subsequence  $T = S \setminus e$  and evaluate the likelihood of the sequence in which the event  $e$  is placed at each possible position in the subsequence  $T$ . I keep updating the sequence  $S$  until convergence. Again I optimise the single cluster sequence  $S$  from different random starting sequences until the algorithm converges to the maximum likelihood solution.

### 7.2.5 Implementation

All experiments were performed using Matlab on a standard workstation (Intel Core i7, 3.1 GHz, 8GB memory). For each of the model fitting procedures (hierarchical,

two-cluster, and single cluster estimation), I repeated the optimisation for 25 starting points to find the maximum likelihood solution, checking that the optimisation displayed good convergence. I found that all start points converge to a solution that is within a  $1 \times 10^{-4}\%$  tolerance level (as a percentage of the maximum likelihood), and within the uncertainty estimated by the uncertainty estimation procedure (described subsequently), meaning that each solution is sufficiently close to the maximum likelihood solution to be used for initialisation of the uncertainty estimation procedure (MCMC algorithm). I ran an MCMC algorithm (as in [98]) to draw samples from the posterior  $P(M|X)$ , taking 1,000,000 MCMC samples and checking that the MCMC trace showed good mixing properties. I initialised the MCMC algorithm at the optimal solution found using the hierarchical estimation procedure so a burn-in period was not required. I further performed 10-fold cross-validation of the model fitting and MCMC results to check that they were reproducible.

The computational complexity of the model fitting varies depending on three factors. The first factor is the time taken to calculate  $P(M|X)$ , which is of order  $O(INJC)$ , where  $I$  is the number of biomarkers,  $N$  is the number of biomarker events,  $J$  is the number of subjects, and  $C$  is the number of clusters. The second factor is the maximum number of clusters being fitted  $C_{\max}$ . The total time taken for the full hierarchical optimisation scales with order  $O(\sum_{c=1}^{C_{\max}} c) \approx O(c^2)$ , as each time there is an additional cluster to estimate the parameters for. The third factor is the time taken for the expectation maximisation and MCMC algorithms to converge. As discussed in Chapter 3, although the space of all possible sequences has a size of  $N!$ , where  $N$  is the number of biomarker events, the maximum pairwise distance between any pair of sequences is  $\frac{N(N-1)}{2}$ . This means that the number of iterations required for convergence of the expectation maximisation algorithm, and for the MCMC chain to sample  $P(M|X)$ , should scale with approximately  $O(N^2C)$ . For the set of biomarker events and number of subjects used here, fitting 5 clusters takes approximately 3 hours in total, and performing MCMC sampling for all 5 models takes around 15 hours.

## 7.3 Results

### 7.3.1 Genetic frontotemporal dementia

#### 7.3.1.1 Dynamic clustering of all mutation carriers

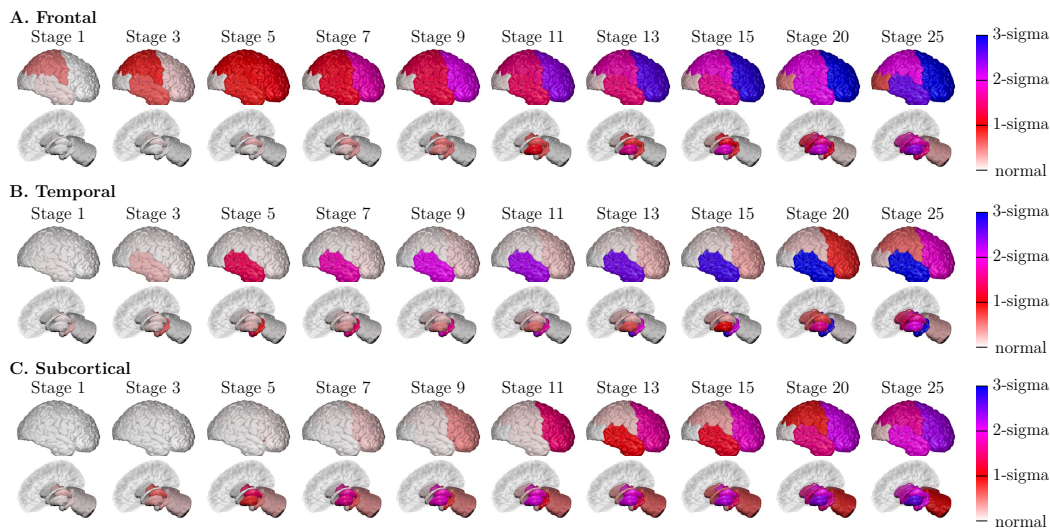
Application of the dynamic clustering model to all mutation carriers in GENFI (see visualisation in Figure 7.1, positional variance diagrams and cross-validation in Supplementary Figure A.1) suggests that there are three population subgroups (the profile of in-sample and out-of-sample model likelihood for different numbers of clusters is shown in Supplementary Figure A.2) with distinct patterns of brain volume loss: a frontal group, a temporal group, and a subcortical group, which account for 51%, 27%, and 22% of the data respectively.

The frontal group (Figure 7.1 A) has a progression pattern that begins with mild (z-score of 1) frontotemporoparietal, cingulate and insula volume loss, followed by aggressive (z-score escalates from 1 to 3 in quick succession) frontal lobe volume loss and moderate volume loss (all approaching a z-score of 2 when frontal lobe volume loss reaches a z-score of 3) in the temporal lobe, parietal lobe, cingulate, insula, putamen and accumbens. There is relatively less volume loss in the temporal lobe, hippocampus, amygdala and insula in this group compared to the temporal group.

The temporal group (Figure 7.1 B) is characterised by volume loss that begins in the hippocampus and amygdala and progresses to the insula, accumbens and temporal lobe, with relative sparing of the frontal and parietal lobe compared to the frontal group. The pattern of volume loss in the temporal group is most aggressive (z-score escalating from 1 to 3 in fastest succession) in the temporal lobe and amygdala, but also progresses quickly in the hippocampus, insula and accumbens.

The subcortical group (Figure 7.1 C) initially have volume loss in the insula, cerebellum, hippocampus, putamen, pallidum and thalamus, which is followed by frontal lobe and accumbens volume loss and then temporoparietal lobe volume loss.

10-fold cross-validation of the dynamic clustering results (Supplementary Figure A.1) shows good agreement across folds.

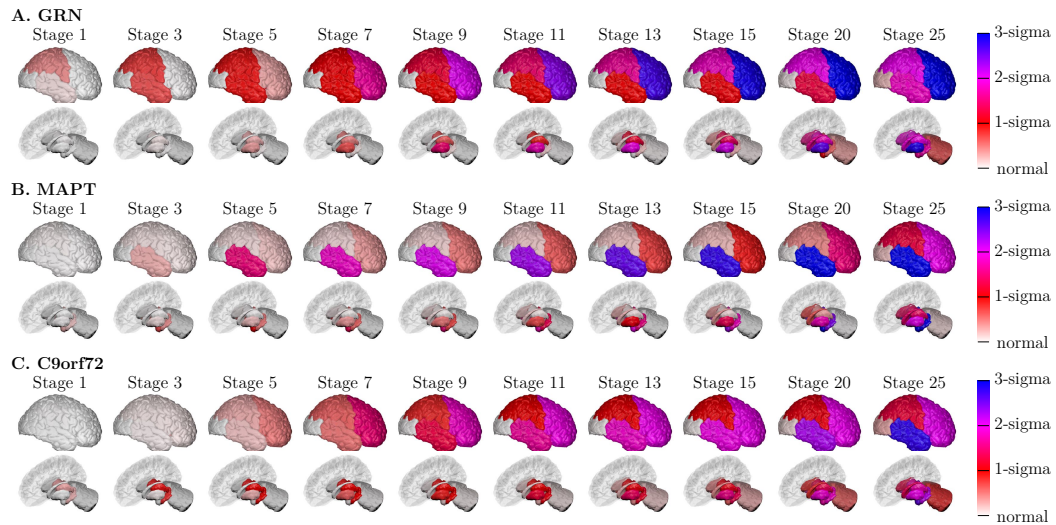


**Figure 7.1:** Dynamic clustering of all mutation carriers in GENFI. Subfigures (A)-(C) show the progression pattern of each of the three dynamic clusters estimated by the model. The cumulative probability each region has reached a particular z-score is shown for different stages along the progression; the cumulative probability of a region going from a z-score of 0-sigma to 1-sigma ranges from 0 in white to 1 in red, the cumulative probability of a region going from a z-score of 1-sigma to 2-sigma ranges from 0 in red to 1 in magenta, and the cumulative probability of a region going from a z-score of 2-sigma to 3-sigma ranges from 0 in magenta to 1 in blue.

### 7.3.1.2 Comparison with fitting a single dynamic cluster to each mutation type

Fitting a single dynamic model to carriers of a mutation in GRN, MAPT and C9orf72 separately (see visualisation in Figure 7.2, positional variance diagrams and cross-validation in Supplementary Figure A.3) reveals a distinct sequence of brain volume loss for each mutation type. The patterns of volume loss in the GRN and MAPT groups (Figure 7.2 A and Figure 7.2 B) map well onto the frontal and temporal patterns of volume loss recovered by dynamic clustering of all mutation carriers (Figure 7.1 A and Figure 7.1 B). However, fitting a single model to the C9orf72 group (Figure 7.2 C) does not map well onto the subcortical group in Figure 7.1 C. 10-fold cross-validation of the dynamic clustering results (Supplementary Figure A.3) for each mutation group shows good agreement across folds.





**Figure 7.2:** Fitting a single dynamic cluster to each mutation type in GENFI separately. Subfigures (A)-(C) show the progression pattern obtained by fitting a single dynamic cluster to each of the three mutation carrier groups separately. Diagrams as in Figure 7.1.

### 7.3.1.3 Dynamic clustering of GRN, MAPT and C9orf72 carrier groups

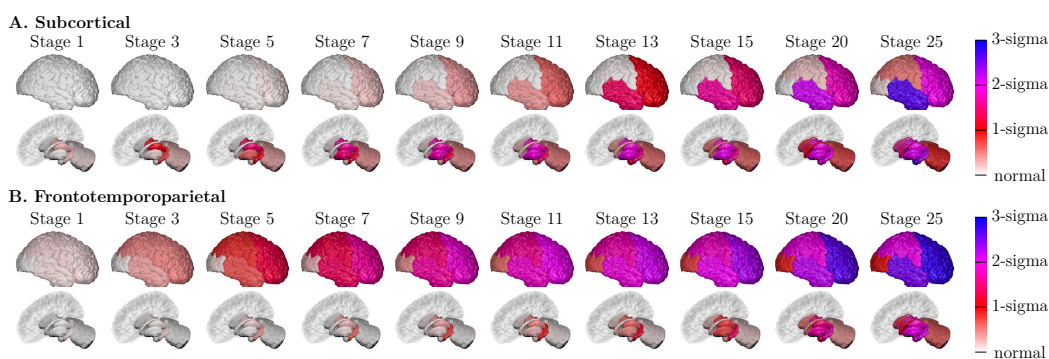
I explored the hypothesis that there may be multiple phenotypes within the GRN, MAPT and C9orf72 carriers by performing dynamic clustering on the GRN, MAPT and C9orf72 mutation carriers respectively.

In the GRN mutation carrier group I found no evidence of there being more than one cluster (Supplementary Figure A.4).

In the MAPT group I found that a two cluster model fitted better than a one cluster model (see Supplementary Figures A.5, A.6 and A.7), with the first cluster being a temporal cluster accounting for 73% of the data, in good agreement with that in Figure 7.1 B and Figure 7.2 B, and the second cluster being a slightly frontotemporoparietal cluster that accounted for 27% of the data. However, the MAPT carrier group has only 41 subjects and this cluster has high uncertainty so it is difficult to conclude much from this second cluster apart from the presence of outliers within the MAPT carrier group.

By performing dynamic clustering of C9orf72 mutation carriers I found that the C9orf72 group are best described by two sequences of brain volume loss (see vi-

sualisation in Figure 7.3, positional variance diagrams and cross-validation in Supplementary Figure A.8, model likelihood in Supplementary Figure A.9): the first cluster (Figure 7.3 A) maps well onto the subcortical group recovered by dynamic clustering of all mutation carriers (Figure 7.1 C), the second cluster (Figure 7.3 B) has a progression pattern that begins with frontotemporoparietal, insula and hippocampal volume loss, and is followed by aggressive frontal and temporal volume loss and moderate volume loss in the cingulate, insula, putamen and accumbens, albeit with high uncertainty in the ordering of the frontal, temporal and parietal volume loss. 10-fold cross-validation of the dynamic clustering results (see Supplementary Figure A.8) for each mutation group shows good agreement across folds. All of the C9orf72 carriers (both affected and unaffected) that are assigned to the subcortical cluster in Figure 7.1 C are assigned to the subcortical cluster in Figure 7.3 A.



**Figure 7.3:** Dynamic clustering of C9orf72 mutation carriers. Subfigures (A)-(B) show the progression pattern of each of the two dynamic clusters estimated by the model. Diagrams as in Figure 7.1.

Of note is that the uncertainty in the sequence of z-score events estimated by the positional variance diagram (Supplementary Figure A.8) encompasses the possibility that the pattern of volume loss in Figure 7.3 B (see positional variance diagram in Supplementary Figure A.8) is the same as the frontal pattern of volume loss in Figure 7.1 A (see positional variance diagram in Supplementary Figure A.1). This means that the cluster in Figure 7.3 B could represent a distinct frontotemporoparietal pattern of volume loss, or a more frontal pattern of volume loss consistent with that in Figure 7.1 A. However, performing a post-hoc analysis of asymmetry in af-

affected mutation carriers assigned to the frontal cluster (Figure 7.1 A) reveals that there are significant differences between affected C9orf72 and GRN carriers assigned to the frontal subtype, with affected GRN carriers having significantly (two sample t-test) more asymmetric atrophy patterns than affected C9orf72 carriers ( $p = 1.12 \times 10^{-5}$ ). Moreover, the asymmetry in the affected C9orf72 carriers assigned to the frontal cluster was not significantly different (two-sample t-test) from affected C9orf72 carriers assigned to the temporal, subcortical or normal-appearing cluster.

I performed several further post-hoc analyses to check for differences between the two C9orf72 groups estimated in Figure 7.3. I classified subjects into frontotemporoparietal, subcortical or normal-appearing groups using the model shown in Figure 7.3, as well as assigning them to a model stage along the progression pattern for their cluster, which corresponds to the average number of z-score events in the sequence that have occurred. The normal-appearing cluster contains subjects assigned to model stage 0, which I added to the model to avoid subjects with normal appearing brain volumes having equal probability of belonging to the initial stage of either of the two groups. I found no statistically significant differences in age (two-sample t-test), estimated years from onset (two sample t-test), model stage (chi-squared test), or scanner field strength (Fisher's exact test), between affected or unaffected C9orf72 mutation carriers assigned to the frontotemporoparietal and subcortical atrophy patterns.

#### 7.3.1.4 Classification of mutation type using dynamic clustering

The frontal, temporal and subcortical patterns of brain volume loss shown in Figure 7.1 can be used to classify subjects. To avoid the situation where subjects with normal appearing brain volumes have equal probability of belonging to the initial stage of any of the three groups, I added in a normal-appearing cluster, which consists of all subjects at stage 0 of the model. Table 7.1 shows the proportion of affected mutation carriers assigned to each dynamic cluster using in-sample models and out-of-sample models obtained from 10-fold cross-validation.

I found that the majority of the affected GRN and MAPT carriers are assigned to the frontal and temporal subtypes respectively (93% of affected GRN carriers

**A.**

	Normal-appearing	Frontal (Figure 7.1 A)	Temporal (Figure 7.1 B)	Subcortical (Figure 7.1 C)
GRN	0% (0)	93% (13)	0% (0)	7% (1)
MAPT	0% (0)	9% (1)	91% (10)	0% (0)
C9orf72	0% (0)	46% (12)	27% (7)	27% (7)

**B.**

	Normal-appearing	Frontal (Figure 7.1 A)	Temporal (Figure 7.1 B)	Subcortical (Figure 7.1 C)
GRN	0% (0)	93% (13)	0% (0)	7% (1)
MAPT	0% (0)	18% (2)	82% (9)	0% (0)
C9orf72	0% (0)	54% (14)	27% (7)	19% (5)

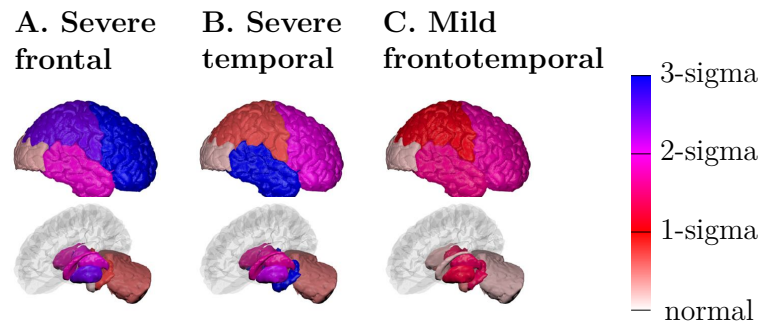
**Table 7.1:** Proportion of affected mutation carriers in GENFI assigned to each dynamic cluster using (A) in-sample models, and (B) out-of-sample models obtained from 10-fold cross-validation. Each entry is the percentage (number) of subjects of a particular mutation type assigned to that cluster.

are assigned to the frontal subtype, 91% of affected MAPT carriers are assigned to the temporal subtype). In the affected C9orf72 mutation carriers however, I found that the majority (46%) are assigned to the frontal subtype, with a further 27% being assigned to each of the temporal and subcortical subtypes. In the unaffected GRN and MAPT carriers I found that the majority (52% and 58% respectively) are assigned to the normal-appearing group. In the unaffected C9orf72 carriers only 24% are assigned to the normal-appearing group, with 39%, 32% and 5% of the unaffected C9orf72 carriers being assigned to the frontal, subcortical and temporal clusters respectively.

### 7.3.1.5 Comparison to static clustering

I ascertained the utility of having a dynamic component to the clustering technique by comparing the results to static clustering, i.e. estimating clusters that have a single constant mean and standard deviation. To ensure a fair comparison between static and dynamic clustering I included a fixed normal-appearing cluster in the static clustering model, with a mean of 0 and a standard deviation of 1 for all biomarkers, as I did for stage 0 of the dynamic clustering. Figure 7.4 visualises the static clusters, Supplementary Table A.1 shows the parameter estimates for the

static clustering, and Table 7.2 shows the proportion of affected mutation carriers assigned to each static cluster using in-sample models and out-of-sample models obtained from 10-fold cross-validation.



**Figure 7.4:** Static clustering of all mutation carriers in GENFI. Subfigures (A)-(C) show the progression pattern of each of the three static clusters estimated by static clustering. Each static cluster consists of a single stage in which each region is coloured according to its z-score value.

**A.**

	Normal-appearing	Severe Frontal (Figure 7.4 A)	Severe Temporal (Figure 7.4 B)	Mild Frontotemporal (Figure 7.4 C)
GRN	0% (0)	64% (9)	0% (0)	36% (5)
MAPT	0% (0)	0% (0)	82% (9)	18% (2)
C9orf72	12% (3)	15% (4)	27% (7)	46% (12)

**B.**

	Normal-appearing	Severe Frontal (Figure 7.4 A)	Severe Temporal (Figure 7.4 B)	Mild Frontotemporal (Figure 7.4 C)
GRN	0% (0)	64% (9)	0% (0)	36% (5)
MAPT	0% (0)	9% (1)	64% (7)	27% (3)
C9orf72	12% (3)	15% (4)	27% (7)	46% (12)

**Table 7.2:** Proportion of affected mutation carriers in GENFI assigned to each static cluster using (A) in-sample models, and (B) out-of-sample models obtained from 10-fold cross-validation. Each entry is the percentage (number) of subjects of a particular mutation type assigned to that cluster.

Static clustering gives three disease clusters (Figure 7.4): a severe frontal cluster, a severe temporal cluster, and a mild frontotemporal cluster. These clusters are not dissimilar from the results of dynamic clustering. However, when I classify sub-

jects into mutation types using static clustering, I see that the third mild frontotemporal cluster can represent an early stage of either the severe frontal or the severe temporal subtype, meaning that it is more difficult to separate out the mutation types using the static clusters. In the affected GRN and MAPT carriers, 64% and 82% of affected carriers are assigned to the frontal and temporal clusters respectively, compared with 93% and 91% for the dynamic clustering. 36% and 18% of the affected GRN and MAPT carriers are assigned to the mild frontotemporal cluster, as the clusters cannot separate the early stages of a frontal or temporal pattern of volume loss from a milder, more diffuse, disease process. As with the dynamic clustering, in the affected C9orf72 group I see heterogeneity in the cluster assignments: 46% are assigned to the frontotemporal cluster, 15% are assigned to the frontal cluster, 27% are assigned to the temporal cluster, and 12% are assigned to the normal-appearing cluster. In the unaffected GRN and MAPT carriers I again find that the majority (80% and 79% respectively) are assigned to the normal-appearing cluster. In the unaffected C9orf72 carriers only 49% are assigned to the normal-appearing cluster, with 49% being assigned to the frontotemporal cluster, and 2% being assigned to the frontal cluster. Note that the high proportion of unaffected GRN and MAPT carriers assigned to the normal-appearing cluster can be achieved using the dynamic clustering model if desired by simply increasing the model stage threshold that is considered as normal-appearing. For example, assigning those with model stage less than 5 to the normal-appearing group results in a proportion of 84% and 83% (compared to 80% and 79% for static clustering) of the unaffected GRN and MAPT carriers being assigned to the normal-appearing cluster, whilst still maintaining 93% and 91% (compared to 64% and 82% for static clustering) of the affected GRN and MAPT carriers being assigned to the frontal and temporal clusters.

### 7.3.2 Alzheimer's disease

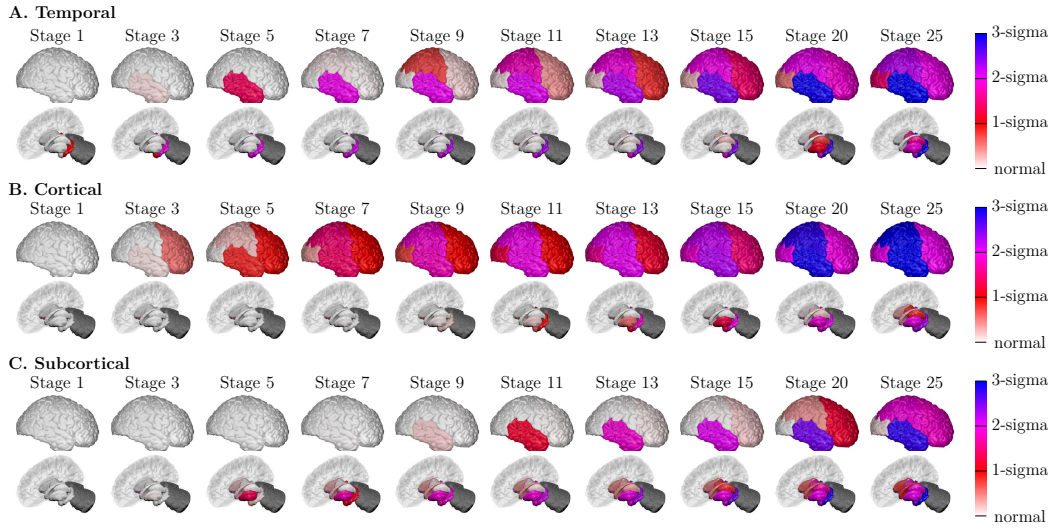
#### 7.3.2.1 Dynamic clustering of the ADNI dataset

I applied the dynamic clustering technique to 3T data from ADNI (see visualisation in Figure 7.5, positional variance diagrams and cross-validation in Supplementary Figure A.10, model likelihood in Supplementary Figure A.11). I found that the 3T

ADNI data is best described as three subtypes (Figure 7.5): a temporal subtype that accounts for 35% of the data, a cortical subtype that accounts for 38% of the data, and a subcortical subtype that accounts for 27% of the data. The temporal subtype in Figure 7.5 A starts with atrophy in the hippocampus, amygdala and temporal lobe, with these three regions also becoming more severe first. The cortical subtype in Figure 7.5 B starts in the accumbens and cingulate and progresses to cortical regions in the frontal, temporal, parietal and occipital lobes. The cortical subtype then goes on to affect the hippocampus, amygdala and putamen, with atrophy in the parietal and temporal lobe atrophy becoming more severe. The subcortical subtype in Figure 7.5 C begins with atrophy in the pallidum, accumbens and putamen, which goes on to affect the hippocampus and amygdala, and then progresses to the temporal lobe. 10-fold cross-validation of the results (Supplementary Figure A.10) shows good agreement across folds. For further validation I repeated the analysis on 1.5T data from ADNI (see Supplementary Figures A.12, A.13 and A.14). Dynamic clustering revealed 4 clusters in the 1.5T dataset, including temporal, cortical and subcortical clusters in good agreement with the 3T data, but also a fourth parietal cluster that accounted for 4% of the data, which is not seen in any of the 3T cluster models. The main characteristic of the parietal cluster was aggressive parietal lobe atrophy, but also severe, but slightly less aggressive, frontal, temporal, and occipital lobe atrophy.

### 7.3.2.2 Association with conversion from mild cognitive impairment to Alzheimer's disease

I tested if model stage and subtype were associated with increased risk of conversion from MCI to AD by fitting a Cox proportional hazards model with time of AD diagnosis as the time to event data, and controlling for age, sex, education and number of APOE4 alleles. Time to event data for subjects who did not convert was considered censored at their last available diagnosis. I found statistically significant associations (Table 7.3) between the risk of conversion from MCI to AD and model stage ( $p = 2.34 \times 10^{-6}$ ), subtype ( $p = 1.09 \times 10^{-5}$ ), and number of APOE4 alleles ( $p = 7.65 \times 10^{-5}$ ) for the 3T dataset. These findings were reproducible in the 1.5T



**Figure 7.5:** Dynamic clustering of 3T ADNI dataset. Subfigures (A)-(C) show the progression pattern of each of the three dynamic clusters estimated by the model. Diagrams as in Figure 7.1. The cerebellum was not included as a region in the ADNI analysis and so is shaded in dark grey.

dataset, where I again found statistically significant associations between the risk of conversion from MCI to AD and model stage ( $p = 1.53 \times 10^{-5}$ ), subtype ( $p = 6.94 \times 10^{-3}$ ), and number of APOE4 alleles ( $p = 1.07 \times 10^{-3}$ ). In both datasets I find that the temporal subtype is associated with the fastest conversion times, followed by the cortical, and then the subcortical subtype.

### 7.3.2.3 Association with decline in cognitive test scores

I tested if model stage and subtype were associated with faster decline in MMSE score in AD subjects by fitting a linear model predicting decline in MMSE from subtype, model stage, MMSE at baseline, age, sex, education and number of APOE4 alleles. I estimated decline in MMSE by fitting a linear model to each individual's longitudinal MMSE scores. I found no statistically significant associations with MMSE decline in the 3T dataset, however in the 1.5T dataset, which has longer follow-up, I found statistically significant associations between decline in MMSE and subtype ( $p = 1.40 \times 10^{-5}$ ), model stage ( $p = 7.19 \times 10^{-3}$ ) and age ( $p = 2.86 \times 10^{-2}$ ). Across both the 1.5T and the 3T datasets the average rate of MMSE decline was fastest in the cortical group, followed by the temporal group, and then the subcortical group.



	Model cluster	Model stage	Age	Sex	Education	APOE4
N-S-C-T	1.60 <sup>‡</sup>	1.10 <sup>‡</sup>	0.98	0.97	0.93	1.76 <sup>‡</sup>
N-S	1.50	1.10 <sup>*</sup>	1.02	1.53	0.86	1.91 <sup>*</sup>
S-C	1.68	1.10 <sup>**</sup>	0.95	1.12	0.94	1.63 <sup>*</sup>
C-T	1.67 <sup>*</sup>	1.10 <sup>†</sup>	0.97	0.81	0.95	1.78 <sup>†</sup>
N-C	3.58 <sup>**</sup>	1.06	0.98	0.80	1.02	1.85 <sup>**</sup>
S-T	2.15 <sup>*</sup>	1.12 <sup>‡</sup>	1.00	1.12	0.85 <sup>**</sup>	1.80 <sup>**</sup>
N-T	2.90 <sup>*</sup>	1.12 <sup>†</sup>	1.02	0.86	0.92	1.97 <sup>†</sup>

**Table 7.3:** Hazards ratios for risk of conversion from MCI to AD in ADNI 3T dataset. Each row shows a different Cox proportional hazards model. For the first model (N-S-C-T) it is assumed that the hazard ratio increases multiplicatively from the Normal Appearing cluster (N) to the Subcortical cluster (S) to the Cortical cluster (C) to the Temporal cluster (T), i.e. the N-S-C-T model predicts that each cluster has a hazards ratio 1.60 times that of the previous cluster. In the remaining models only two groups are compared at a time to remove the assumption that there is the same multiplicative increase between each consecutive cluster. Each column shows the estimated hazard ratio for that variable. Statistical significance is indicated as: <sup>\*</sup> =  $p < 0.05$ , <sup>\*\*</sup> =  $p < 0.01$ , <sup>†</sup> =  $p < 1 \times 10^{-3}$ , <sup>‡</sup> =  $p < 1 \times 10^{-4}$ .

## 7.4 Discussion

I have demonstrated the use of a data-driven model to uncover disease subtypes with distinct patterns of biomarker evolution in genetic FTD and AD. I have shown the ability of the model to recover genetic subgroups of FTD without prior knowledge of their mutation type. The model provides good separation of GRN and MAPT groups, however, in the C9orf72 group the model predicts that there are two population subgroups with distinct progression patterns. This result provides new insights into the disease mechanisms related to the C9orf72 mutation. In AD, the model predicts that there are three population subgroups with distinct progression patterns: a temporal pattern, a cortical pattern and a subcortical pattern. Moreover, I found that these subgroups had distinct characteristics at longitudinal follow-up, with the temporal subtype being associated with the fastest conversion from MCI to AD, and the cortical group being associated with the fastest decline in cognitive test scores.

### 7.4.1 Genetic frontotemporal dementia

#### 7.4.1.1 Dynamic clustering of all mutation carriers

Performing dynamic clustering of all mutation carriers reveals that there are three distinct patterns of volume loss: a frontal pattern, a temporal pattern and a subcortical pattern (Figure 7.1). Comparing these clusters to the clusters obtained by fitting a single dynamic cluster to each of the mutation types separately (Figure 7.2) shows that the model can recover the frontal pattern of the GRN group (Figure 7.1 A similar to Figure 7.2 A) and the temporal pattern of the MAPT group (Figure 7.1 B similar to Figure 7.2 B).

In the C9orf72 group as a whole, however, the progression pattern does not align well with the third subcortical cluster predicted by the model (Figure 7.1 C different to Figure 7.2 C). Performing dynamic clustering of each of the mutation carrier groups individually provides insight into the problem by revealing that the C9orf72 group are in fact best modelled by two clusters (Figure 7.3): one with a subcortical pattern and one with a frontotemporoparietal pattern. The subcortical pattern in Figure 7.3 A is in agreement with the subcortical pattern found in all mutation carriers in Figure 7.1 C.

There is still one cluster that is not recovered by the model: the frontotemporoparietal pattern in Figure 7.3 B. This is likely because the frontotemporoparietal cluster (Figure 7.3 B) is similar to the frontal cluster (Figure 7.1 A), and so does not improve the model likelihood enough for inclusion in the model. This hypothesis is supported by the fact that increasing the number of clusters from three to four leads to the addition of a frontotemporoparietal cluster (data not shown). The hypothesis is further supported by previous findings, which have shown that the differences between the neuroimaging signatures of C9orf72 carriers and GRN carriers are small when asymmetry is not taken into account [179], although previous work has only considered the C9orf72 group as a whole, rather than as two groups with distinct progression patterns. It has been shown that GRN mutation carriers have highly asymmetric atrophy patterns in comparison to C9orf72 mutation carriers. Including a measure of asymmetry would therefore likely improve the model's ability to sepa-

rate the frontal cluster (Figure 7.1 A) and the frontotemporoparietal cluster (Figure 7.3 B).

#### 7.4.1.2 Dynamic clustering of GRN mutation carriers

I found that the GRN mutation carriers are best modelled by a single progression pattern: the frontal pattern shown in Figures 7.1 A and 7.2 A. This result is in agreement with previous studies, which have found that GRN mutations are associated with the most severe frontal and parietal lobe atrophy [179], with relative sparing of the medial temporal lobe.

#### 7.4.1.3 Dynamic clustering of MAPT mutation carriers

The general pattern of atrophy in the MAPT carriers (Figures 7.1 B and 7.2 B) is in agreement with previous findings, showing severe temporal lobe atrophy, with relative sparing of the frontal and parietal lobes [179]. By performing dynamic clustering of the MAPT mutation carriers, I found that the MAPT group are best modelled as two groups with distinct progression patterns (Supplementary Figures A.5 and A.6), with 73% of the MAPT carriers following a temporal progression pattern (which shows good correspondence with the general pattern of atrophy shown in Figures 7.1 B and 7.2 B), and 26% following a frontotemporoparietal pattern, somewhat similar to that of the C9orf72 subgroup in Figure 7.3 B, but with very high uncertainty (Supplementary Figure A.6 B). This uncertainty is likely due to the small number of MAPT carriers (35, of which 24 are unaffected and 11 are affected), but could also represent variation in the alternative progression pattern. Whilst this result does suggest that there is heterogeneity within the MAPT group, the uncertainty is too high to make any conclusions about the ordering within the second MAPT subgroup.

#### 7.4.1.4 Dynamic clustering of C9orf72 mutation carriers

In the C9orf72 group I found that there are two distinct progression patterns (Figure 7.3): a subcortical pattern and a frontotemporoparietal pattern. This result is well aligned with previous work, which suggests that the C9orf72 mutation carriers are the most heterogeneous [179, 189], but in contrast to previous work I am

able to characterise this heterogeneity as two distinct progression patterns. The subcortical progression pattern (Figure 7.3 A) is unique to the C9orf72 group, and resembles previous findings, which have shown significant subcortical involvement in C9orf72 mutation carriers [189]. The frontotemporoparietal pattern (Figure 7.3 B) is similar to the progression pattern of the GRN group (Figure 7.2 A), which is interesting as previous studies have shown that it is difficult to differentiate between C9orf72 and GRN mutation carriers [179, 189] without accounting for asymmetry. A post-hoc analysis of asymmetry revealed that the GRN group had significantly more asymmetric atrophy than the C9orf72 mutation carriers that were assigned to the frontotemporoparietal pattern. This finding suggests that although the affected regions are similar, the underlying mechanisms in the GRN and C9orf72 groups are different, with the mechanism in the GRN mutation carriers preferentially affecting one side of the brain.

#### 7.4.1.5 Classification of mutation type using dynamic clustering

I demonstrated that the model can differentiate between the GRN and MAPT mutation groups (Table 7.1), with 93% of affected GRN carriers being assigned to the frontal subtype (Figure 7.1 A) and 91% of affected MAPT carriers being assigned to the temporal subtype (Figure 7.1 B). In the affected C9orf72 mutation carriers however, I found that the majority (46%) of participants were assigned to the frontal pattern in Figure 7.1 A, with a further 27% being assigned to the temporal pattern (Figure 7.1 B), and 27% to the subcortical pattern (Figure 7.1 C). Including a measure of asymmetry would improve the ability to differentiate between the GRN mutation carriers and the subset of the C9orf72 mutation carriers that are assigned to the frontal pattern of volume loss in Figure 7.1 A.

#### 7.4.1.6 Comparison to static clustering

I compared the dynamic clustering model to a standard ‘static’ clustering model, in which each cluster constitutes a static biomarker profile, rather than a temporal progression pattern. Compared to static clustering, dynamic clustering provides a much more detailed picture of disease progression, and better separates the different genetic frontotemporal dementia subtypes. This is because dynamic clustering

is able to separate disease subtypes from disease stages, whereas static clustering produces clusters that are a mixture of disease subtypes and disease stages.

## 7.4.2 Alzheimer's disease

### 7.4.2.1 Dynamic clustering of ADNI dataset

Dynamic clustering of the ADNI dataset reveals that there are three clusters with distinct patterns of biomarker evolution: a temporal, a cortical and a subcortical pattern. These clusters are seen in both the 1.5T and the 3T datasets. These three clusters are somewhat in agreement with previous neuropathological findings [27], which have defined three subtypes of AD based on the distribution of neurofibrillary tangles: typical, hippocampal-sparing and limbic-predominant. The temporal cluster (Figure 7.5 A and Supplementary Figure A.12 A) resembles the pattern of atrophy seen in typical AD; the cortical cluster (Figure 7.5 B and Supplementary Figure A.12 B) is similar to hippocampal-sparing AD; the subcortical cluster reflects the atrophy pattern of limbic-predominant AD (Figure 7.5 C and Supplementary Figure A.12 C). These three distinct patterns of atrophy have been shown on MRI previously [154, 190], both when groups are defined pathologically [154], or in a more data-driven manner [190].

In contrast to previous studies, I am able to characterise the earliest sites of regional volume loss for each subgroup. I find that regional volume loss in the temporal group (typical AD, Figure 7.5 A and Supplementary Figure A.12 A) is first detectable in the hippocampus and amygdala, whereas regional volume loss in the cortical group (hippocampal-sparing AD, Figure 7.5 B and Supplementary Figure A.12 B) is first detectable in the insula and cingulate, and regional volume loss in the subcortical group (limbic-predominant AD, Figure 7.5 C and Supplementary Figure A.12 C) is first detectable in the pallidum and putamen.

When using the 3T data I additionally found that the accumbens area is implicated as one of the earliest detectable sites of regional volume loss in the cortical and subcortical groups, which I did not find in the 1.5T data. This might be because small areas, such as the accumbens, are more visible in the 3T data, but this result requires further validation.

In the 1.5T dataset I found an additional cluster with a parietal atrophy pattern, accounting for 4% of the data. This cluster may represent outliers who have posterior cortical atrophy, which is an atypical variant of AD that causes atrophy of the posterior part of the cerebral cortex, leading to progressive disruption of visual processing. Parietal atrophy patterns were also found as a possible fourth cluster by the study of Zhang et al. [190], although their study found that optimal model of sporadic AD consisted of three clusters.

#### 7.4.2.2 Association with conversion from mild cognitive impairment to Alzheimer's disease

I found that each of the dynamic clusters of AD was associated with different patient outcomes. I found significant associations between the risk of conversion from MCI to AD and subtype in both the 3T and 1.5T datasets. Moreover, I also found significant associations with model stage and risk of conversion, meaning that the staging and subtyping information provided by the model could provide complementary information for both patient staging and stratification.

#### 7.4.2.3 Association with decline in cognitive test scores

In the 1.5T dataset, I found statistically significant associations between decline in MMSE and subtype ( $p = 1.40 \times 10^{-5}$ ), model stage ( $p = 7.19 \times 10^{-3}$ ) and age ( $p = 2.86 \times 10^{-2}$ ). Across both the 1.5T and the 3T datasets the average rate of MMSE decline was fastest in the cortical group, followed by the temporal group, and then the subcortical group. These results are in agreement with the study of Zhang et al. [190], which also found that cortical atrophy patterns are associated with the fastest decline in MMSE score, followed by temporal atrophy patterns and then subcortical atrophy patterns.

## 7.5 Conclusion

I have developed a dynamic clustering model that characterises disease subtypes with distinct biomarker trajectories (contribution 1.3.3 (a), Chapter 1, page 41). I have demonstrated the ability of the model to recover known genetic subgroups of

FTD without prior knowledge of their mutation type for GRN and MAPT mutation carriers (contribution 1.3.3 (b), Chapter 1, page 41). The model further provides new insights into heterogeneity in the C9orf72 mutation carrier group by revealing that this group are best modelled by two disease progression patterns: a subcortical pattern and a frontotemporoparietal pattern. In AD, dynamic clustering predicts there are three longitudinal patterns of atrophy: temporal, cortical and subcortical (contribution 1.3.3 (c), Chapter 1, page 41). I further demonstrated that these patterns are associated with distinct characteristics at longitudinal follow-up. The dynamic clustering model presented here has wide potential further applications: for characterising heterogeneity in other diseases, and as a patient staging and stratification mechanism for precision medicine.





## **Chapter 8**

# **Further work**

## **8.1 Overview**

This chapter discusses three broad areas of opportunity for future work: (1) further applications of the models developed in this thesis; (2) opportunities for technological enhancement of the models presented in this thesis; (3) methodological advances that can be made in the wider spectrum of data-driven models.

## **8.2 Further applications**

### **8.2.1 Application in neurodegenerative diseases**

There are numerous opportunities for further application of the models developed in this thesis in various neurodegenerative diseases.

#### **8.2.1.1 Alzheimer's disease**

In AD, this thesis has explored the sequence of biomarker changes in both the sporadic and dominantly-inherited disease forms. Chapter 3 investigated the sequence in which a multi-modal set of biomarkers become abnormal in sporadic AD. Chapter 4 performed a similar study using dominantly-inherited AD biomarker data. Chapter 7 examined the heterogeneity of biomarker trajectories of regional volume loss in sporadic AD.

The model developed in Chapter 7 was only applied to volumetric MRI data here. In future it will be interesting to characterise the heterogeneity in the sequence of biomarker changes for a more multi-modal set of biomarkers that includes CSF

measures, PIB-PET, FDG-PET, and cognitive test scores. Of particular interest is the inclusion of cognitive test scores as the subtypes found in Chapter 7 were shown to have different rates of decline in cognitive test scores. The incorporation of subscales that measure more specific types of cognitive deficits will determine whether there are observable differences in the sequence in which different areas of cognition decline for each subtype.

In Chapter 4 I observed that there were subtle differences in the sequence of biomarker abnormality in dominantly-inherited AD for different mutation types (APP vs. PSEN1 and PSEN2) and genetic subgroups (APOE4 positive vs. APOE4 negative). Application of the dynamic clustering model developed in Chapter 7 will determine whether the dominantly-inherited AD data is better described by multiple trajectories. If so, it will be interesting to see whether the optimal clustering of dominantly-inherited AD biomarker trajectories corresponds to different mutation types or genetic subgroups, or whether the relationship is more nuanced.

### 8.2.1.2 Frontotemporal dementia

In Chapter 7 I developed a dynamic clustering model that finds population subgroups with distinct sequences of biomarker changes. Application of the model to genetic frontotemporal dementia (FTD) revealed that the participants carrying a C9orf72 mutation are best described by two patterns of regional MRI volume loss: a subcortical and a frontotemporoparietal pattern.

Further analysis of the C9orf72 group including a more multi-modal set of biomarker data will provide interesting insights into the underlying disease biology of these two observed patterns. Of particular interest is the inclusion of neuropsychological test scores and subscales. It is thought that subcortical involvement in the C9orf72 group causes the neuropsychiatric symptoms that a subset of the C9orf72 mutation carriers present with [189]. A subset of C9orf72 mutation carriers present with behavioural variant FTD [189]; this group may be better aligned with the frontotemporoparietal pattern of volume loss.

Another opportunity for future work is the clustering of sporadic FTD, in which patients exhibit different patterns of presentation, to uncover data-driven sub-

groups with distinct patterns of biomarker changes. These data-driven sporadic FTD subgroups can then be compared with the genetic subgroups established in Chapter 7.

### 8.2.1.3 Other neurodegenerative diseases

The models proposed in this thesis are equally applicable to other neurodegenerative diseases. Determining the sequence of biomarker changes in other neurodegenerative diseases, such as Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, vascular dementia and posterior cortical atrophy will allow a direct comparison of the progression of the different diseases and produce a set of models that can potentially be used for differential diagnosis.

## 8.2.2 Application to other diseases or processes

The models developed in this thesis potentially generalise to a wider range of diseases and processes. Examples include chronic obstructive pulmonary disease (COPD), cancer, multiple sclerosis, normal ageing and developmental processes. Different diseases can present specific methodological challenges however, for example in multiple sclerosis there are both destructive and healing processes, which violates the assumption typically made by data-driven models of monotonic progression patterns.

## 8.3 Methodological developments of the models presented in this thesis

This section discusses methodological developments that are more specific to the models presented in this thesis.

### 8.3.1 Event distributions

All of the models presented here depend on the definition of a distribution of 'normal', and, in the case of the event-based models (EBMs) developed in Chapters 3, 4 and 6, 'abnormal' measurements. However, often the exact definition of these distributions is unclear. For example, the control population may have small numbers or contain outliers, or it may be difficult to decide what constitutes 'abnormal', par-

ticularly for biomarkers that change slowly over time, which are not well described as a binary transition from ‘normal’ to ‘abnormal’. Future work will explore modelling the uncertainty in these distributions to establish the effect of the choice of event distribution parameters on the resulting models of biomarker evolution. Another key area for further development is the incorporation of non-gaussian event distributions, this is particularly important for cognitive tests, which are typically discrete measures.

### 8.3.2 Partial sequences

The models presented in this thesis assume that all biomarker events in the sequence occur. In the case of the EBMs developed in Chapters 3, 4 and 6 this is particularly problematic as when a biomarker has no signal the distribution of ‘normal’ and ‘abnormal’ measurements may overlap. This can lead to the corresponding event erroneously appearing early in the sequence. Development of models with partial event sequences could resolve this problem. In such a model only biomarkers that are better modelled as two distributions: ‘normal’ and ‘abnormal’ would be included as events.

With the z-score model developed in Chapter 7 this problem is resolved as the disease group will have a similar distribution of z-scores to the control group, and therefore no change in z-score will be observed. However, it is still desirable to develop models that do not assume all biomarker events occur in order to find the most parsimonious model of disease progression. In the case of fitting multiple biomarker trajectories for example, it may be that a specific sub-population, for example a set of misdiagnosed patients, is only observed early in the disease because later in the disease it is easier to filter out those who have been misdiagnosed. In this case it would be nicer to be able to truncate the set of biomarker events, rather than assuming that the other events will occur eventually.

### 8.3.3 Time

A more fundamental limitation of the models proposed in this thesis is that they do not incorporate time: they are temporal in the sense that the events happen in an

order, but there is no notion of event duration. In some cases this is desirable as it allows the models to be fitted to purely cross-sectional datasets, but the majority of datasets do have some short-term follow-up time points available. When longitudinal follow-up is available it should be possible to determine how long it takes for each biomarker to transition from one state to the next by observing patients in which this transition occurs. Future work will develop a mathematical formulation of this idea to allow the time between each transition to be determined.

### 8.3.4 Missing data

In Chapter 4 a simple adaptation of the EBM for use with missing data was developed. This adaptation was based upon the assumption that the data is missing at random and is not applicable to the z-score models developed in Chapter 7. Future work will develop more advanced techniques for handling missing data that alleviate this assumption, for example using data imputation strategies.

### 8.3.5 High-dimensional biomarkers

In their current form the proposed models take as input scalar biomarker values. This means that complex high-dimensional biomarkers, such as those from imaging, have to be pre-processed to derive scalar measurements, such as regional volumes. Extension of the models for use with high-dimensional biomarkers would alleviate this problem and further provide much more spatially detailed pictures of disease progression. There are two major methodological barriers to developing such a model. The first is the combinatorial complexity of the problem. The models work by searching for a sequence of biomarker changes, the space of possible sequences is  $N!$ , where  $N$  is the number of biomarker changes being modelled. This complexity becomes intractable for high-dimensional biomarkers. Future work might develop techniques that reduce the dimensionality of this search space. The second problem is that high-dimensional biomarkers often have complex correlation structures that are not modelled here. The recent work of Bilgel et al. [133] proposed a set of correlation structures for images, which it may be possible to incorporate into the models.

### **8.3.6 Within-subject models**

The models developed in this thesis compare biomarker measurements cross-sectionally between subjects. In order to develop models that are optimal for application to longitudinal within-subject data, further adaptations of the models will need to be made that can estimate intra-subject variability in biomarker measurements.

## **8.4 Broader technical advances**

In this section I discuss technical advances that can be made to a broader range of data-driven models.

### **8.4.1 Mixed pathology**

Data-driven models to date have modelled a single disease process, or a set of individual disease processes, where each patient belongs to a single disease. However, neuropathological studies show that the majority of dementia cases have mixed pathology. Models that can disentangle different pathological processes therefore present an interesting opportunity for further methodological development.

### **8.4.2 Integration of data-driven and mechanistic models**

There are several types of mechanistic model that have provided interesting insights into neurodegenerative disease progression but depend on a priori disease staging information. Future work could integrate these models with data-driven statistical methodology to remove their reliance on a priori knowledge of disease stage. This will allow more complex disease progression patterns to be recovered.

Network models describe disease progression as evolving along structural or functional networks, starting from a particular region. To infer the region from which the disease process is initiated, they compare the patterns of atrophy predicted by network models with different starting regions to the pattern of atrophy seen in AD [125]. The choice of starting region is therefore dependent only on how well the network model aligns with a coarse end stage pattern of atrophy, and more complex disease mechanisms cannot be estimated. Using data-driven mod-

els, more detailed pictures of disease progression can be recovered for the full disease time course. Incorporating these data-driven progression patterns into network modelling approaches could allow more complex models to be fitted. For example, current network models are typically based on the prion hypothesis, which posits that disease proteins are physically transmitted between neurons. The alternative hypothesis is that of selective vulnerability: that some regions are more vulnerable to disease pathology than others. It has been proposed that there may be a combination of these processes in operation [54]. Data-driven network models may be able to determine whether one or both of these processes better models disease progression patterns.

Spatiotemporal models describe the spatial as well as the temporal evolution of disease progression. However, these models are dependent on a priori knowledge of disease stage and so their temporal resolution is coarse. The development of data-driven spatiotemporal models offers the possibility of combining the detailed spatial pictures of disease progression recovered by spatiotemporal models with the fine-grained temporal resolution of data-driven models.

### **8.4.3 Dimensionality reduction**

Another potential development of data-driven models is the inclusion of dimensionality reduction techniques to take complex data types, such as images, and extract key information relevant to disease progression at each stage. Sparse learning techniques, for example, could be used to choose a subset of voxels of an MRI image that are the most sensitive to disease progression at each stage. A particular advantage of combining sparse learning techniques with data-driven models would be that at any point in time you could model only the voxels that were relevant to that particular stage. For example, brain regions that become abnormal late in the disease would only be modelled at the late disease stages.

### **8.4.4 Incorporating additional data types**

To date, data-driven models have largely considered biomarker measurements alone. The incorporation of additional data types such as genetics, demographic

information, lifestyle factors or medical records can potentially make the models more powerful. Such measurements might be used to reduce heterogeneity in the biomarker measurements, thereby increasing the power for detecting disease pathology compared to a control population. For instance, it has been shown that common genetic variants have specific influences on the volume of subcortical brain structures [191]. By having a model of the expected volume of different brain structures based on a subjects genetic profile, it may be possible to detect more subtle disease-related changes in volume. Additional data types could also be used to provide supporting information for a particular diagnosis or staging assignment, for example medical records may be indicative of an increased risk for a particular disease.

#### **8.4.5 Validation**

The simulation system presented in Chapter 5 provides a basic framework for generating synthetic data to validate data-driven disease progression models. This framework is used to test the robustness of two data-driven models: the EBM and a differential equation model. Future work will compare a larger range of data-driven models, and develop more complex simulation systems that can simulate MRI scans to validate high-dimensional models. Another important aspect of the validation of data-driven models is testing their reproducibility, and determining whether they can be translated for use in new datasets.

#### **8.4.6 Patient staging systems**

A promising outcome of data-driven models is the natural patient staging system they provide: subjects can be matched to their most probable point along the disease trajectory according to their biomarker measurements. However, these models are not specifically optimised for patient staging. Further work will explore the development of data-driven models that are designed to make optimal patient staging and diagnostic decisions. There are several open problems in this area. One aspect is biomarker utility (discussed in detail in the subsequent subsection), i.e. determining the most useful set of measurements to take in order to make a particular staging or diagnostic decision. Another aspect is designing an intuitive staging system for



clinical practice: a rich continuous model may provide more information than is necessary, for example, confusing the diagnostic process.

#### **8.4.7 Biomarker utility**

Complex models can provide biological insights by integrating information from multimodal biomarkers to build up a detailed picture of the progression of different pathologies in relation to one another. For application of a biomarker in clinic however, the cost of taking different types of measurement needs to be considered. To date there has been relatively little work on optimising the set of measurements that are made to inform a particular clinical decision. Data-driven models can provide a basis on which such an optimisation can be performed. For example, the fine-grained quantitative models of biomarker evolution generated using data-driven models can be used to determine the sensitivity of various biomarkers for measuring changes in disease stage at a particular point along the disease time course. New mathematical models of biomarker utility can be developed that weight this sensitivity against the cost of acquiring each measurement to decide which biomarker, or set of biomarkers, is optimal. Similar technical developments can be made that inform differential diagnosis.

### **8.5 Summary**

The models proposed in this thesis have wide potential application to a range of diseases and developmental processes. There are several possible methodological enhancements to the models presented here, of particular interest are the incorporation of time measures and extension to high dimensional biomarkers. Data-driven models in general are an emerging technology with a broad range of opportunities for further work, such as the incorporation of additional data types, optimal design of patient staging systems, and determination of biomarker utility.



## Chapter 9

# Summary and conclusion

### 9.1 Chapter 1. Introduction

AD is a progressive neurodegenerative disorder that is characterised by the accumulation of amyloid plaques and neurofibrillary tangles in brain tissue. These pathologies are thought to give rise to downstream neurodegeneration and cognitive deficits, however the biological mechanisms of AD are not well understood. Biomarkers have been developed to allow the pathologies of AD to be monitored in vivo. These biomarkers include CSF measures of  $A\beta_{1-42}$ , phosphorylated tau and total tau, volumetric MRI measures of neurodegeneration and cognitive test scores.

Understanding the quantitative evolution of biomarkers in AD is key to providing precision medicine, which will enable more effective clinical trials, as well as personalised treatment plans once disease-modifying drugs are available. In clinical trials, quantitative models of biomarker progression can be used to identify and monitor the presymptomatic disease stages, during which treatments may be more effective. Additionally, a quantitative picture of biomarker progression can provide insights into the underlying disease biology by, for example, indicating which is the initiating disease pathology, or elucidating interactions between different pathogenic processes. However, reconstructing a detailed quantitative picture of biomarker progression in AD is difficult due to the coarse disease staging measures, which typically consist of just three stages, and the long disease time course, which is thought to span several decades.

This thesis developed data-driven models of biomarker progression that characterise the quantitative evolution of biomarker measurements without reliance on a priori knowledge of disease stage. The models developed in this thesis provide a fine-grained picture of biomarker evolution, and have clinical utility for patient staging.

## **9.2 Chapter 2. State of the art in Alzheimer's disease progression modelling**

Various different progression models have been applied to AD. Neuropathological models and animal models provide interesting biological insights but do not directly provide quantitative measures of disease stage, limiting their utility for patient staging and monitoring. Biomarker models have been developed to allow subjects to be monitored in vivo. Scalar biomarker models combine information from multiple single dimensional biomarkers. These models are highly relevant clinically as they integrate measurements from well established biomarkers, however their dependence on a priori knowledge of disease stage limits their temporal resolution. High dimensional models learn new biomarkers from complex data types, such as images, but still require subjects to be indexed by disease stage. Data-driven models [99] develop novel statistical methodology to allow the temporal progression of biomarkers to be reconstructed without prior knowledge of each individual's position along the disease time course. Such models can recover more detailed disease progression patterns for fine-grained patient staging. However, data-driven models are still an emerging technology, and require further validation and refinement before they are translated into a useful clinical tool.

## **9.3 Chapter 3. A data-driven model of biomarker changes in sporadic Alzheimer's disease**

In this chapter I demonstrated the use of a probabilistic generative model to explore the biomarker changes occurring as AD develops and progresses [134]. I enhanced the event-based model (EBM) [98] for use with a multi-modal sporadic disease data

set (contribution 1.3.1 (a), Chapter 1, page 39). This allowed the sequence in which AD biomarkers become abnormal to be determined without reliance on a-priori clinical diagnostic information or explicit biomarker cut points. The model also characterises the uncertainty in the ordering and provides a natural patient staging system.

I used the EBM to determine the sequence of biomarker abnormality and its uncertainty in various population subgroups from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (contribution 1.3.1 (b), Chapter 1, page 39). I used patient stages assigned by the EBM to discriminate CN from AD subjects, and predict conversion from MCI to AD and CN to MCI.

The model predicted that CSF levels become abnormal first, followed by rates of atrophy, then cognitive test scores, and finally regional brain volumes. In amyloid positive or APOE4 positive subjects, the model predicted with high confidence that the CSF biomarkers become abnormal in a distinct sequence:  $A\beta_{1-42}$ , p-tau, t-tau. However, in the broader population t-tau and p-tau were found to be earlier CSF markers than  $A\beta_{1-42}$ , albeit with more uncertainty. The model's staging system strongly separated CN and AD subjects, and predicted conversion from MCI to AD, and from CN to MCI. By fitting Cox Proportional Hazards models, I found that baseline model stage was a significant risk factor for conversion from both MCI to AD and CN to MCI.

The results support hypothetical models of biomarker ordering in amyloid positive and APOE4 positive subjects, but suggest that biomarker ordering in the wider population may diverge from this sequence. The model provides useful disease staging information across the full spectrum of disease progression, from CN to MCI to AD.

## **9.4 Chapter 4. A data-driven model of biomarker changes in dominantly-inherited Alzheimer's disease**

This chapter explored the sequence in which biomarker changes occur in dominantly-inherited AD. The EBM [98] was enhanced for use with missing data (contribution 1.3.1 (c), Chapter 1, page 39), facilitating its application to a multi-modal dominantly-inherited AD dataset from the Dominantly Inherited Alzheimer Network (DIAN) study.

I estimated the sequence of biomarker abnormality for various population subgroups of the DIAN study (contribution 1.3.1 (d), Chapter 1, page 39). I found that the sequence of biomarker abnormality for all mutation carriers broadly agrees with previous work: PIB-PET deposition becomes abnormal first, followed by CSF levels, and then regional volumetric MRI measures, cognitive test scores and FDG-PET hypometabolism. In comparison to previous work, the sequence estimated by the EBM does not depend on familial age of onset and provides more detailed progression patterns than have been seen previously. Although based on small numbers, the results for the population subgroups suggest minor subtle differences in the ordering for different genetic groups. In particular, CSF  $A\beta_{1-42}$  becomes abnormal before CSF tau in the APP mutation carriers and APOE4-positive group, with the reverse for the PSEN1 and PSEN2 mutation carriers and APOE4-negative group. I further demonstrated the utility of the EBM for patient staging in dominantly-inherited AD: the model separated non-carriers from affected mutation carriers with a high classification accuracy and showed good longitudinal consistency at follow-up.

## **9.5 Chapter 5. A simulation system for biomarker evolution in neurodegenerative disease**

In this chapter I presented a framework for simulating cross-sectional or longitudinal biomarker data sets from neurodegenerative disease cohorts that reflect the

temporal evolution of the disease and population diversity [167] (contribution 1.3.2 (a), Chapter 1, page 40). The simulation system provides a mechanism for evaluating the performance of data-driven models of disease progression, which bring together biomarker measurements from large cross-sectional (or short term longitudinal) cohorts to recover the average population-wide dynamics.

I demonstrated the use of the simulation framework in two different ways (contribution 1.3.2 (b), Chapter 1, page 40). First, to evaluate the performance of the EBM for recovering biomarker abnormality orderings from cross-sectional datasets. Second, to evaluate the performance of a differential equation model (DEM) for recovering biomarker abnormality trajectories from short-term longitudinal datasets.

The results highlighted several important considerations when applying data-driven models to sporadic disease datasets as well as key areas for future work. The system revealed several important insights into the behaviour of each model. For example, the EBM is robust to noise on the underlying biomarker trajectory parameters, under-sampling of the underlying disease time course and outliers who follow alternative event sequences. However, the EBM is sensitive to accurate estimation of the distribution of normal and abnormal biomarker measurements. In contrast, I found that the DEM is sensitive to noise on the biomarker trajectory parameters, resulting in an over estimation of the time taken for biomarker trajectories to go from normal to abnormal. This over estimate is approximately twice as long as the actual transition time of the trajectory for the expected noise level in neurodegenerative disease datasets.

This simulation framework is equally applicable to a range of other models and longitudinal analysis techniques.

## **9.6 Chapter 6. Multiple orderings of events in disease progression**

The EBM [98] relies on the assumption that all subjects follow a single event sequence. This is a major simplification for sporadic disease data sets, which are highly heterogeneous, include distinct subgroups, and contain significant propor-

tions of outliers.

In this chapter I relaxed this assumption by considering two extensions to the EBM [170]: a generalised Mallows model, which allows subjects to deviate from the main event sequence, and a Dirichlet process mixture of generalised Mallows models, which models clusters of subjects that follow different event sequences, each of which has a corresponding variance. I developed a Gibbs sampling technique to infer the parameters of the two models from multi-modal biomarker data sets.

I applied this technique to data from ADNI to determine the sequence in which brain regions become abnormal in sporadic AD, as well as the heterogeneity of that sequence in the cohort. I found that the generalised Mallows model estimates a larger variation in the event sequence across subjects than the original EBM. Fitting a Dirichlet process model detected three subgroups of the population with different event sequences. The Gibbs sampler additionally provided an estimate of the uncertainty in each of the model parameters, for example an individual's latent disease stage and cluster assignment.

A major limitation of this work is that the models are overly complex. In particular, the variance parameter required for the generalised Mallows model hugely increases the model complexity without any gain in the clinical utility of the models. In the subsequent chapter I developed a more parsimonious model of heterogeneous biomarker progression patterns.

## **9.7 Chapter 7. A data-driven model of disease subtypes with distinct patterns of biomarker evolution in frontotemporal dementia and Alzheimer's disease**

Neurodegenerative diseases are often heterogeneous between and within pathologies, with both different types (between) and the same type (within) of pathology spreading with different spatial patterns. In frontotemporal dementia (FTD) differ-



ent between and within pathology subtypes have been described. In AD three within pathology subtypes have been defined according to the distribution of neurofibrillary tangle counts. However, current knowledge of neurodegenerative disease subtypes is limited to a coarse overall pattern of volume loss in groups of subjects at a single disease stage; little consideration has been given to the temporal progression that leads to that pattern.

In this chapter I presented a probabilistic generative model of disease progression that uncovers population subgroups with distinct patterns of biomarker evolution (contribution 1.3.3 (a), Chapter 1, page 41). I demonstrated the ability of the model to recover known genetic subtypes of FTD using volumetric MRI data from the Genetic Frontotemporal Dementia Initiative (GENFI) (contribution 1.3.3 (b), Chapter 1, page 41). The results further revealed that subjects with a pathogenic mutation in chromosome 9 open reading frame 72 (C9orf72) are best described by two disease progression patterns: a subcortical and a frontotemporoparietal pattern.

Application of the model to AD using volumetric MRI data from ADNI uncovered three data-driven subtypes: temporal, cortical, and subcortical, comparable to those observed in neuropathological studies (contribution 1.3.3 (c), Chapter 1, page 41). In comparison to previous studies, the model characterises the evolution of these subtypes as the disease progresses, including the earliest sites of regional volume loss. By fitting a Cox proportional hazards model I found that the time taken to convert from MCI to AD is significantly different between the different subtypes, independently of model stage, with the temporal subtype being associated with the fastest conversion times.

This approach has the potential to facilitate precision medicine by uncovering disease subtypes and quantifying their corresponding patterns of biomarker evolution across diseases.

## **9.8 Chapter 8. Further work**

In Chapter 8 I discussed potential opportunities for further work, covering three main areas: (1) further applications of the models developed in this thesis; (2) op-

portunities for technological enhancement of the models presented in this thesis; (3) methodological advances that can be made in the wider spectrum of data-driven models.

## 9.9 Conclusion

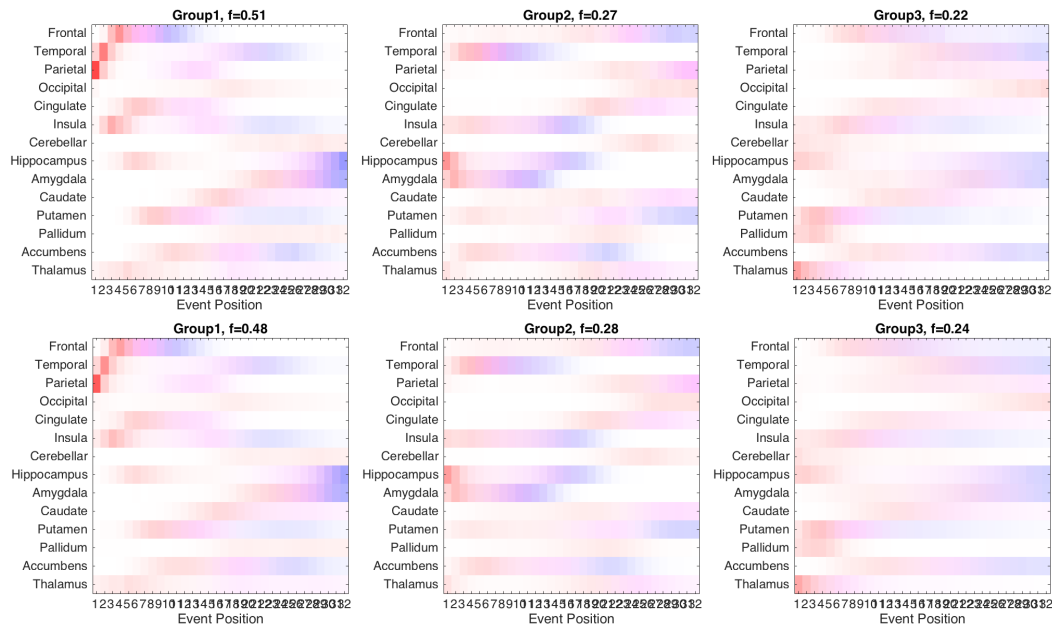
This thesis has explored the progression and heterogeneity of AD by developing quantitative models of biomarker evolution. By proposing two adaptations to the EBM [98], I have been able to characterise the sequence of biomarker changes in both sporadic and dominantly-inherited AD. In sporadic AD the models provide support for hypothetical models of disease progression [74], without reliance on a priori knowledge of disease stage or the use of cut points defining abnormal biomarker levels. I further demonstrated the ability of the model to stage patients throughout the full disease time course. In dominantly-inherited AD, I found that the series of biomarker changes predicted by the models broadly agrees with current knowledge, but provides a much more detailed picture of disease progression than previous work. I additionally found that there may be subtle differences between the PSEN1, PSEN2 and APP carrier groups. I then developed a simulation system to validate data-driven models of disease progression and applied this system to perform a stability analysis of the EBM and a DEM. Finally, I presented a dynamic clustering technique that reveals population subgroups with distinct patterns of biomarker evolution. I demonstrated the ability of this model to recover known subtypes of genetic FTD. The model further provided interesting biological insights into the C9orf72 mutation carrier group by revealing that this group are best modelled by two progression patterns: a subcortical and a frontotemporoparietal pattern. Dynamic clustering of sporadic AD uncovered three data-driven disease subtypes: temporal, cortical and subcortical, which were found to have different conversion times between diagnoses. The models proposed in this thesis have made several novel contributions to AD research and have wide potential further application to other diseases and developmental processes.

## Appendix A

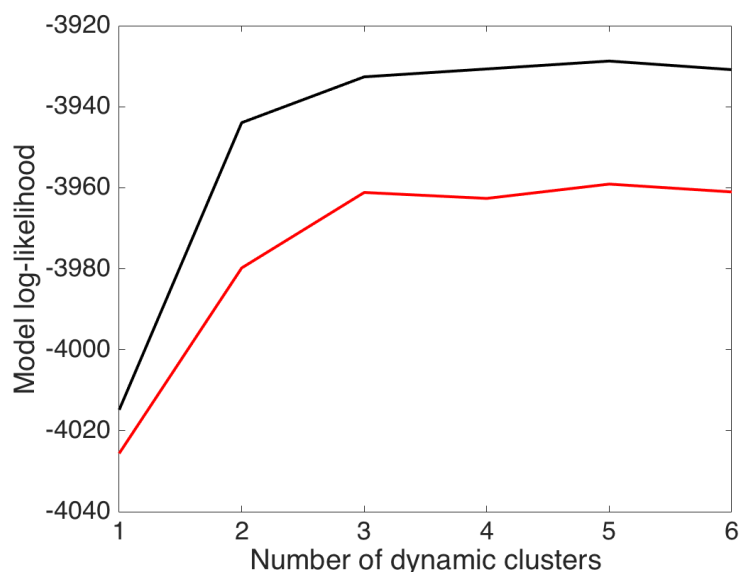
# Supplementary figures and tables for Chapter 7

	Normal- appearing	Severe Frontal	Severe Temporal	Mild Frontotemporal
Frontal	0.00 (1.00)	4.31 (1.00)	1.87 (1.13)	1.44 (1.06)
Temporal	0.00 (1.00)	1.74 (1.17)	4.43 (1.48)	1.29 (1.15)
Parietal	0.00 (1.00)	2.77 (0.88)	0.89 (0.94)	1.02 (0.88)
Occipital	0.00 (1.00)	0.49 (1.19)	0.38 (0.80)	0.39 (1.07)
Cingulate	0.00 (1.00)	2.67 (0.50)	1.35 (0.96)	0.70 (1.01)
Insula	0.00 (1.00)	2.36 (0.79)	3.81 (0.89)	1.40 (1.01)
Cerebellar	0.00 (1.00)	0.79 (0.74)	0.71 (1.03)	0.35 (0.80)
Hippocampus	0.00 (1.00)	0.92 (0.81)	3.48 (1.26)	1.14 (0.69)
Amygdala	0.00 (1.00)	0.25 (0.88)	4.30 (1.38)	0.68 (0.81)
Caudate	0.00 (1.00)	1.60 (0.88)	1.43 (0.88)	0.37 (1.16)
Putamen	0.00 (1.00)	2.74 (0.80)	2.27 (0.84)	1.08 (1.15)
Pallidum	0.00 (1.00)	1.14 (1.11)	1.18 (0.81)	0.53 (1.37)
Accumbens	0.00 (1.00)	1.85 (0.69)	3.15 (0.85)	0.72 (0.98)
Thalamus	0.00 (1.00)	1.83 (0.62)	1.47 (0.82)	1.15 (1.19)

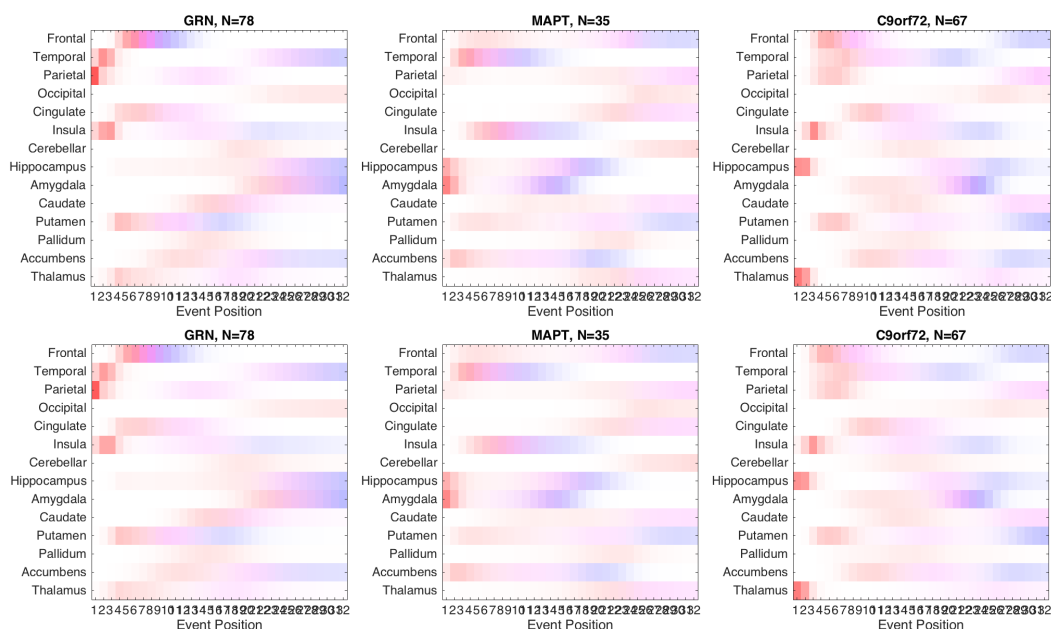
**Table A.1:** Parameters for each of the static clusters shown in Figure 7.4. Each entry is the mean (standard deviation) of the static cluster expressed as a z-score relative to controls (all non-carriers).



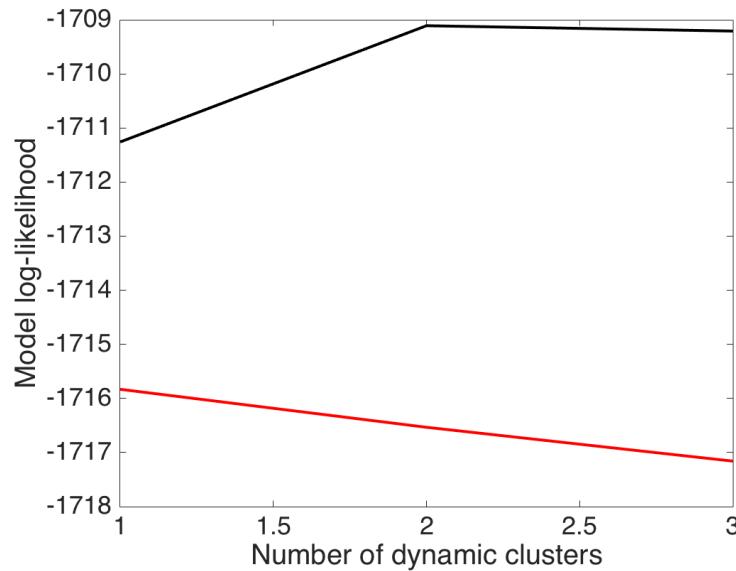
**Figure A.1:** Dynamic clustering of all mutation carriers in GENFI: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. Same result as Figure 7.1, but displayed as positional variance diagrams showing the order of z-score events in each dynamic cluster and its uncertainty from MCMC sampling (top row), and 10-fold cross-validation (bottom row). Each entry in the positional variance diagram represents the proportion of MCMC samples (top row), or the average of the MCMC samples across the 10 cross-validation folds (bottom row), in which z-score events appear at a particular position in the sequence (x-axis). This proportion ranges from 0 in white to 1 in colour. Each z-score is represented as a different colour:  $z=1$  in red,  $z=2$  in magenta, and  $z=3$  in blue. The y-axis keeps the order of events fixed to allow easier comparison across clusters. Where rows have a single coloured block, such as the  $z=1$  parietal lobe event in red in the group 1 positional variance diagram, the ordering is strong and permutations of those events are unlikely.



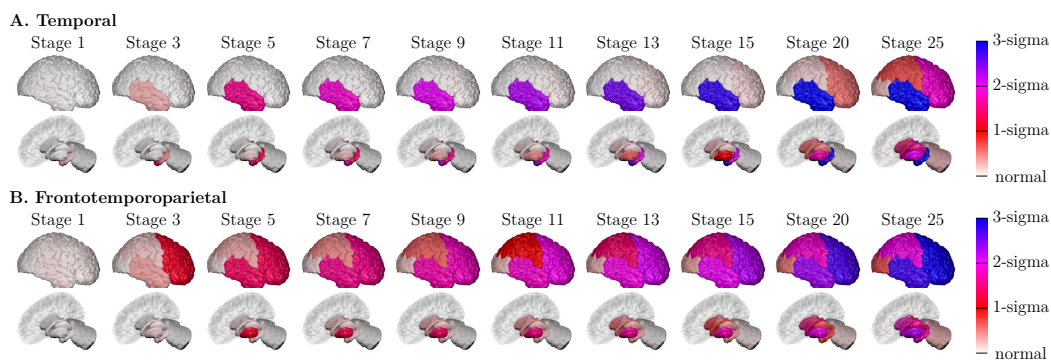
**Figure A.2:** Model likelihood for each of the dynamic cluster models fitted to all mutation carriers in GENFI. In black is the in-sample model likelihood evaluated on the whole dataset, in red is the out-of-sample model likelihood estimated using 10-fold cross-validation. The out-of-sample model likelihood is used to choose the appropriate number of clusters.



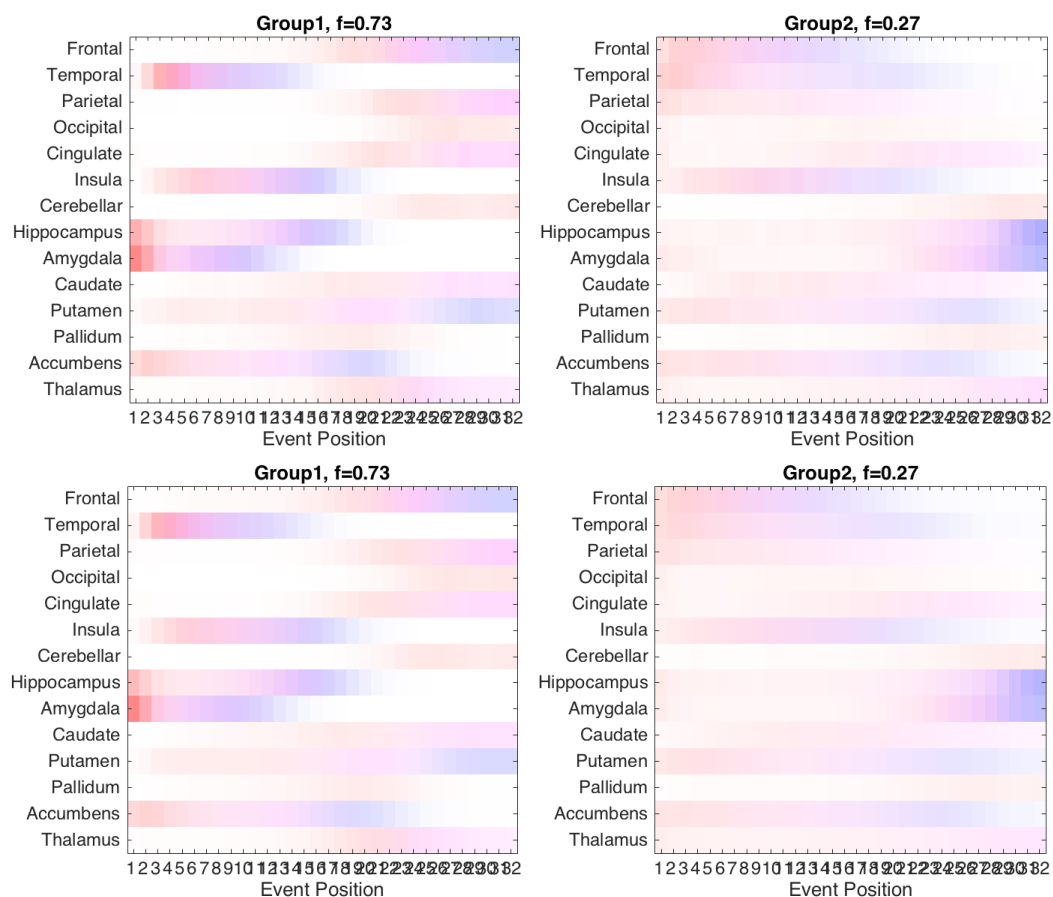
**Figure A.3:** Fitting a single dynamic cluster to each mutation type separately: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. Same result as Figure 7.2, but displayed as positional variance diagrams. Positional variance diagrams as in Figure A.1.



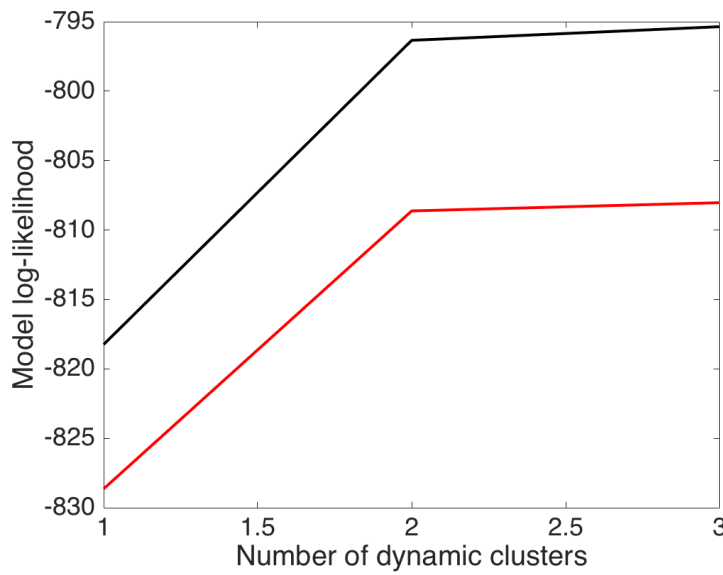
**Figure A.4:** As Figure A.2, but for GRN mutation carriers.



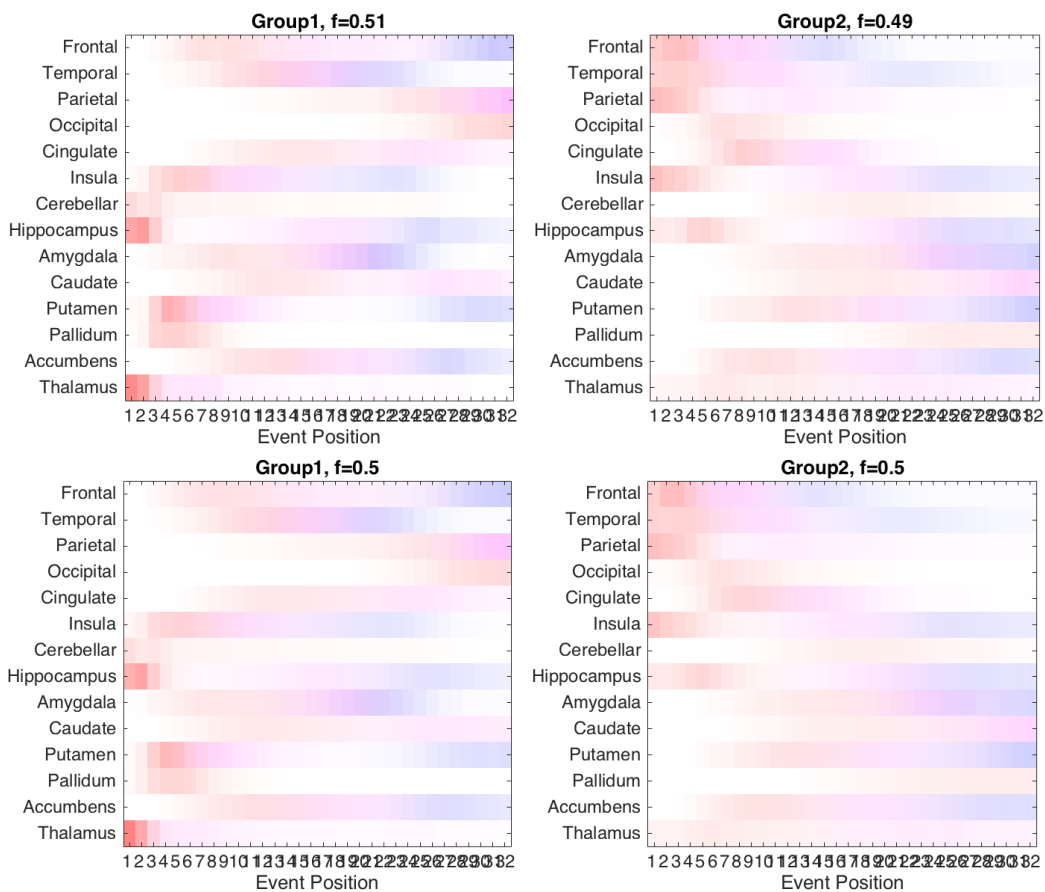
**Figure A.5:** Dynamic clustering of MAPT mutation carriers. Subfigures (A)-(B) show the progression pattern of each of the two dynamic clusters estimated by the model. Diagrams as in Figure 7.1.



**Figure A.6:** Dynamic clustering of MAPT mutation carriers: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. Same result as Figure A.5, but displayed as positional variance diagrams. Positional variance diagrams as in Figure A.1.

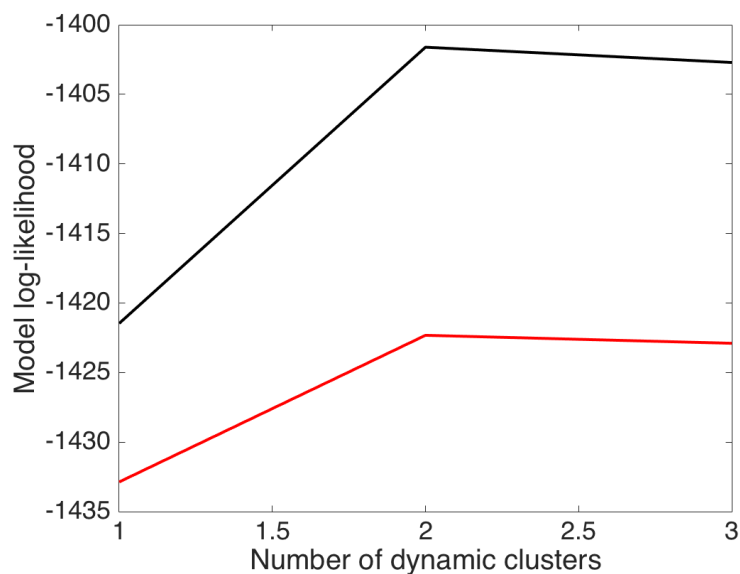


**Figure A.7:** As Figure A.2, but for MAPT mutation carriers.

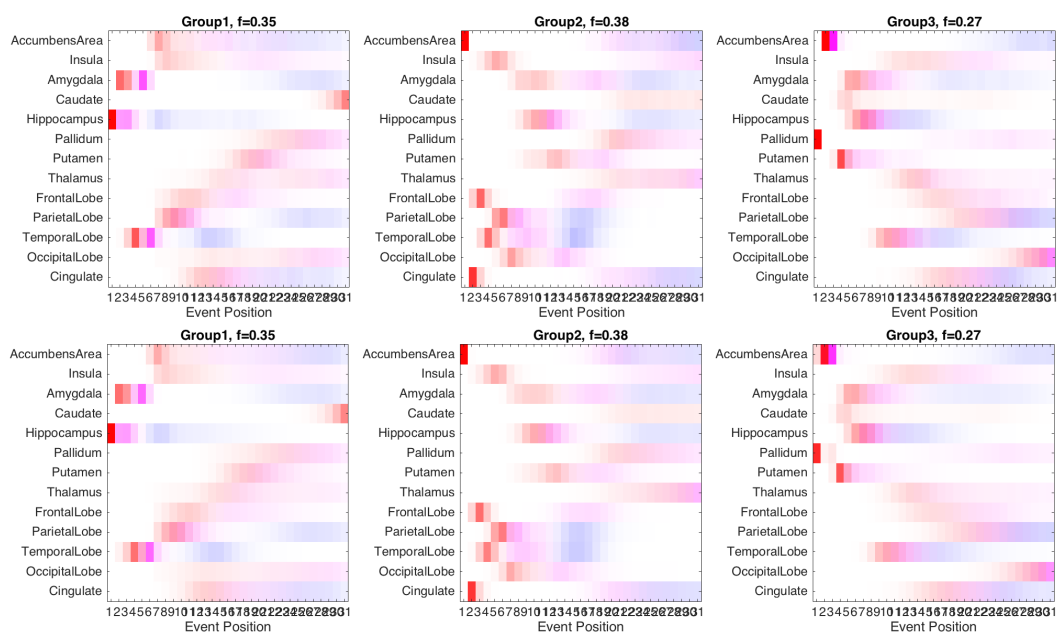


**Figure A.8:** Dynamic clustering of C9orf72 mutation carriers: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. Same result as Figure 7.3, but displayed as positional variance diagrams. Positional variance diagrams as in Figure A.1.

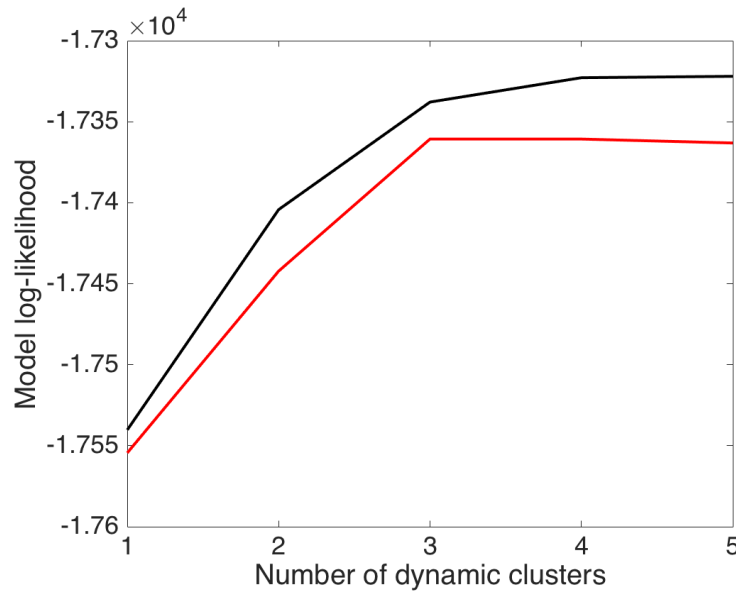




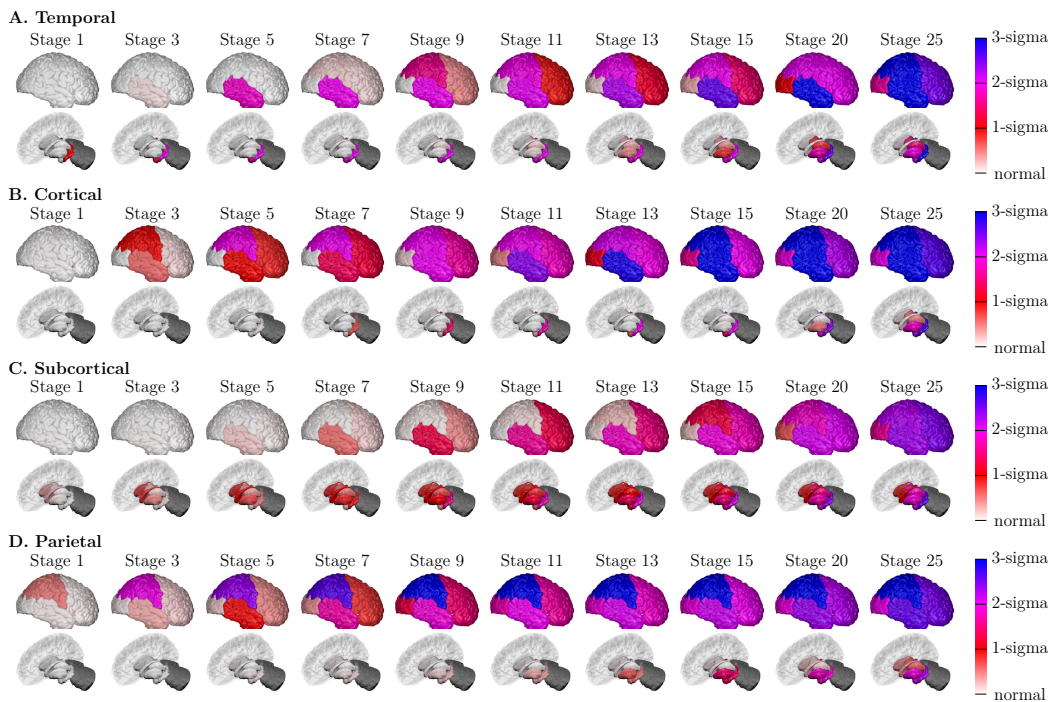
**Figure A.9:** As Figure A.2, but for C9orf72 mutation carriers.



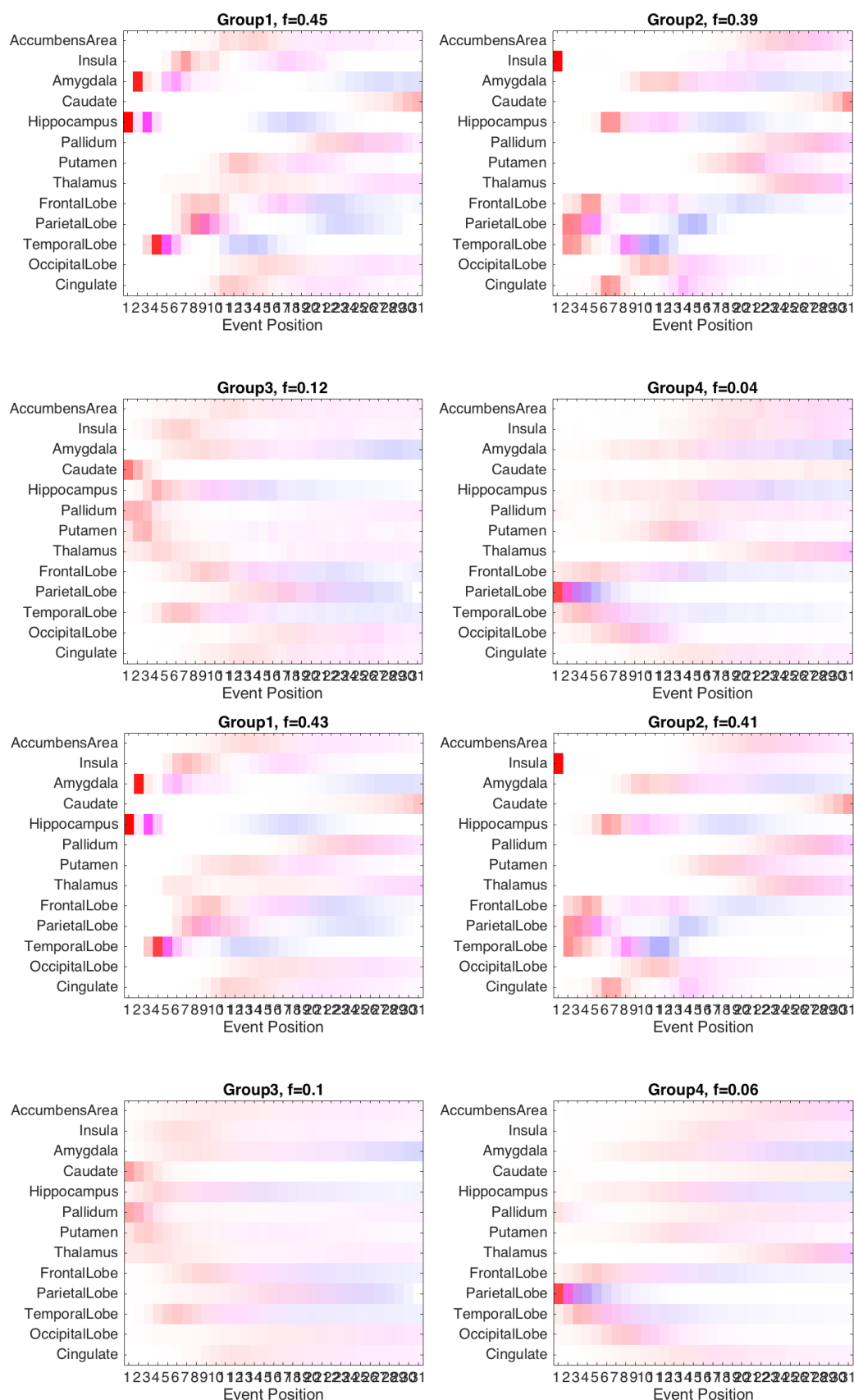
**Figure A.10:** Dynamic clustering of 3T ADNI dataset: MCMC samples of uncertainty on top row, 10-fold cross-validation on bottom row. Same result as Figure 7.5, but displayed as positional variance diagrams. Positional variance diagrams as in Figure A.1.



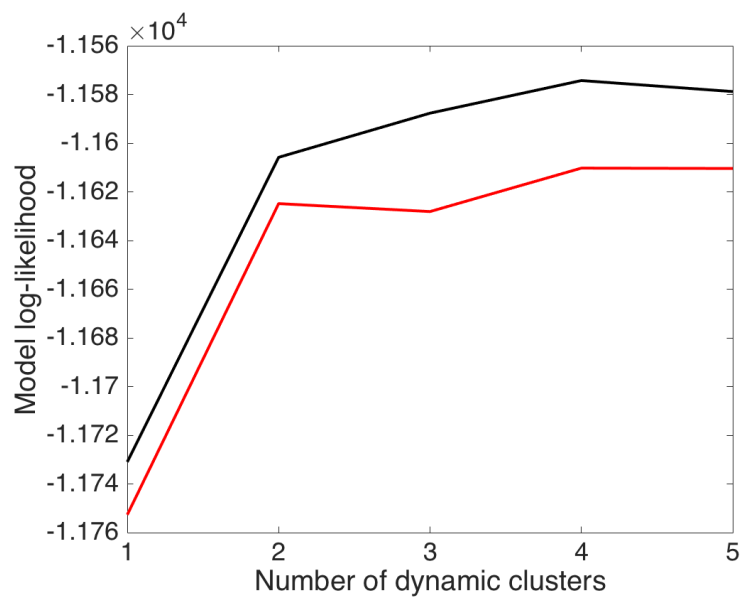
**Figure A.11:** As Figure A.2, but for 3T ADNI dataset.



**Figure A.12:** Dynamic clustering of 1.5T ADNI dataset. Subfigures (A)-(D) show the progression pattern of each of the four dynamic clusters estimated by the model. Diagrams as in Figure 7.1. The cerebellum was not included as a region in the ADNI analysis and so is shaded in dark grey.



**Figure A.13:** Dynamic clustering of 1.5T ADNI dataset: MCMC samples of uncertainty on top two rows, 10-fold cross-validation on bottom two rows. Same result as Figure A.12, but displayed as positional variance diagrams. Positional variance diagrams as in Figure A.1.



**Figure A.14:** As Figure A.2, but for 1.5T ADNI dataset.

# Bibliography

- [1] Alzheimer's Association. Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 12(4):459–509, 2016.
- [2] W. W. Barker, C. A. Luis, A. Kashuba, M. Luis, D. G. Harwood, D. Loewenstein, C. Waters, P. Jimison, E. Shepherd, S. Sevush, N. Graff-Radford, D. Newland, M. Todd, B. Miller, M. Gold, K. Heilman, L. Doty, I. Goodman, B. Robinson, G. Pearl, D. Dickson, and R. Duara. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Disease and Associated Disorders*, 16(4):203–212, 2002.
- [3] R. S. Wilson, E. Segawa, P. A. Boyle, S. E. Anagnos, L. P. Hize, and D. A. Bennett. The Natural History of Cognitive Decline in Alzheimer's Disease. *Psychology and Aging*, 27(4):1008–1017, 2012.
- [4] M. Prince, A. Wimo, M. Guerchet, A. Gemma-Claire, Y.-T. Wu, and M. Prina. World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends. *Alzheimer's Disease International*, 2015.
- [5] A. Alzheimer. Über einen eigenartigen schweren Erkrankungsprozeß der Hirnrinde. *Neurologisches Centralblatt*, 23:1129–36, 1906.
- [6] K. Maurer, S. Volk, and H. Gerbaldo. Auguste D and Alzheimer's disease. *The Lancet*, 349(9064):1546–1549, 1997.

- [7] A. Alzheimer. Über eine eigenartige Erkrankung der Hirnrinde. *Allgemeine Zeitschrift für Psychiatrie und Psychisch-Gerichtliche Medizin*, 64:146–48, 1907.
- [8] E. Kraepelin. Psychiatrie: Ein Lehrbuch für Studierende und Ärzte. *Leipzig: Barth*, pages 593–632, 1910.
- [9] G. Blessed, B. E. Tomlinson, and M. Roth. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *The British Journal of Psychiatry*, 114:797–811, 1968.
- [10] D. Selkoe, E. Mandelkow, and D. Holtzman. Deciphering Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(1):a011460, 2012.
- [11] F. Boller and M. M. Forbes. History of dementia and dementia in history: An overview. *Journal of the Neurological Sciences*, 158(2):125–133, 1998.
- [12] R. Katzman. The prevalence and malignancy of Alzheimer disease: A major killer. *Archives of Neurology*, 33(4):217–218, 1976.
- [13] D. J. Selkoe and J. Hardy. The amyloid hypothesis of Alzheimer’s disease at 25 years. *EMBO Molecular Medicine*, 8(6):1–14, 2016.
- [14] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman. Neuropathological alterations in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, 1(1):a006189, 2011.
- [15] S. Itagaki, P. L. McGeer, H. Akiyama, S. Zhu, and D. Selkoe. Relationship of microglia and astrocytes to amyloid deposits of Alzheimer disease. *Journal of Neuroimmunology*, 24(3):173–182, 1989.
- [16] E. Masliah. Mechanisms of synaptic dysfunction in Alzheimer’s disease. *Histology and Histopathology*, 10(2):509–519, 1995.
- [17] C. J. Pike, B. J. Cummings, and C. W. Cotman. Early association of reactive astrocytes with senile plaques in Alzheimer’s disease. *Experimental Neurology*, 132(2):172–179, 1995.

- [18] R. B. Knowles, C. Wyart, S. V. Buldyrev, L. Cruz, B. Urbanc, M. E. Hasselmo, H. E. Stanley, and B. T. Hyman. Plaque-induced neurite abnormalities: Implications for disruption of neural networks in Alzheimer's disease. *Proceedings of the National Academy of Sciences*, 96(9):5274–5279, 1999.
- [19] B. T. Hyman, C. H. Phelps, T. G. Beach, E. H. Bigio, N. J. Cairns, M. C. Carrillo, D. W. Dickson, C. Duyckaerts, M. P. Frosch, E. Masliah, S. S. Mirra, P. T. Nelson, J. Schneider, D. R. Thal, B. Thies, J. Q. Trojanowski, H. V. Vinters, and T. J. Montine. National institute on aging-Alzheimer's association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia*, 8:1–13, 2012.
- [20] S. M. Gentleman, C. Bruton, D. Allsop, S. J. Lewis, J. M. Polak, and G. W. Roberts. A demonstration of the advantages of immunostaining in the quantification of amyloid plaque deposits. *Histochemistry*, 92(4):355–358, 1989.
- [21] J. C. Morris, M. Storandt, D. W. McKeel Jr., E. H. Rubin, J. L. Price, E. A. Grant, and L. Berg. Cerebral amyloid deposition and diffuse plaques in “normal” aging: Evidence for presymptomatic and very mild Alzheimer's disease. *Neurology*, 46(3):707–719, 1996.
- [22] D. S. Wolf, M. Gearing, D. A. Snowdon, H. Mori, W. R. Markesbery, and S. S. Mirra. Progression of regional neuropathology in Alzheimer disease and normal elderly: Findings from the Nun Study. *Alzheimer Disease and Associated Disorders*, 13(4):226–231, 1999.
- [23] D. R. Thal, U. Rüb, M. Orantes, and H. Braak. Phases of A $\beta$ -deposition in the human brain and its relevance for the development of AD. *Neurology*, 58(12):1791–800, 2002.
- [24] F. Braak, H. Braak, and E. M. Mandelkow. A sequence of cytoskeleton changes related to the formation of neurofibrillary tangles and neuropil threads. *Acta Neuropathologica*, 87(6):554–567, 1994.

- [25] J. C. Augustinack, A. Schneider, E.-M. Mandelkow, and B. T. Hyman. Specific tau phosphorylation sites correlate with severity of neuronal cytopathology in Alzheimer's disease. *Acta Neuropathologica*, 103(1):26–35, 2002.
- [26] H. Braak and E. Braak. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991.
- [27] M. E. Murray, N. R. Graff-Radford, O. A. Ross, R. C. Petersen, R. Duara, and D. W. Dickson. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: A retrospective study. *The Lancet Neurology*, 10(9):785–796, 2011.
- [28] S. E. Arnold, B. T. Hyman, J. Flory, A. R. Damasio, and G. W. Van Hoesen. The topographical and neuroanatomic distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cerebral Cortex*, 1:103–116, 1991.
- [29] P. V. Arriagada, J. H. Growdon, E. T. Hedley-Whyte, and B. T. Hyman. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology*, 42(3):631–639, 1992.
- [30] P. V. Arriagada, K. Marzloff, and B. T. Hyman. Distribution of Alzheimer-type pathologic changes in nondemented elderly individuals matches the pattern in Alzheimer's disease. *Neurology*, 42(9):1681–1688, 1992.
- [31] D. Perl. Neuropathology of Alzheimer's Disease. *Mount Sinai Journal of Medicine*, 77:32–42, 2010.
- [32] R. E. Tanzi. The genetics of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, 2(10):a006296, 2012.
- [33] A. Goate, M.-C. Chartier-Harlin, M. Mullan, J. Brown, F. Crawford, L. Fidani, L. Giuffra, A. Haynes, N. Irving, L. James, R. Mant, P. Newton, K. Rooke, P. Roques, C. Talbot, M. Pericak-Vance, A. Roses, R. Williamson, M. Rossor, M. Owen, and J. Hardy. Segregation of a missense mutation in the



- amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, 349(6311):704–706, 1991.
- [34] R. Sherrington, E. I. Rogaev, Y. Liang, E. A. Rogaeva, G. Levesque, M. Ikeda, H. Chi, C. Lin, G. Li, K. Holman, T. Tsuda, L. Mar, J.-F. Foncin, A. C. Bruni, M. P. Montesi, S. Sorbi, I. Rainero, L. Pinessi, L. Nee, I. Chumakov, D. Pollen, A. Brookes, P. Sanseau, R. J. Polinsky, W. Wasco, H. A. R. Da Silva, J. L. Haines, M. A. Pericak-Vance, R. E. Tanzi, A. D. Roses, P. E. Fraser, J. M. Rommens, and P. H. St George-Hyslop. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*, 375(6534):754–760, 1995.
- [35] E. Levy-Lahad, W. Wasco, P. Poorkaj, D. M. Romano, J. Oshima, W. H. Pettingell, C.-E. Yu, P. D. Jondro, S. D. Schmidt, K. Wang, A. C. Crowley, Y.-H. Fu, S. Y. Guenette, D. Galas, E. Nemens, E. M. Wijsman, T. D. Bird, G. D. Schellenberg, and R. E. Tanzi. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science*, 269(5226):973–977, 1995.
- [36] E. I. Rogaev, R. Sherrington, E. A. Rogaeva, G. Levesque, M. Ikeda, Y. Liang, H. Chi, C. Lin, K. Holman, T. Tsuda, L. Mar, S. Sorbi, B. Nacmias, S. Piacentini, L. Amaducci, I. Chumakov, D. Cohen, L. Lannfelt, P. E. Fraser, J. M. Rommens, and P. George-Hyslop. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature*, 376(6543):775–778, 1995.
- [37] D. Scheuner, C. Eckman, M. Jensen, X. Song, M. Citron, N. Suzuki, T. D. Bird, J. Hardy, M. Hutton, W. Kukull, E. Larson, E. Levy-Lahad, M. Viitanen, E. Peskind, P. Poorkaj, G. Schellenberg, R. Tanzi, W. Wasco, L. Lannfelt, D. Selkoe, and S. Younkin. Secreted amyloid  $\beta$ -protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the pre-senilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Medicine*, 2(8):864–870, 1996.

- [38] W. J. Strittmatter, A. M. Saunders, D. Schmechel, M. Pericak-Vance, J. Englund, G. S. Salvesen, and A. D. Roses. Apolipoprotein E: High-avidity binding to  $\beta$ -amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences*, 90(5):1977–1981, 1993.
- [39] E. H. Corder, A. M. Saunders, N. J. Risch, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell Jr., J. B. Rimmler, P. A. Locke, P. M. Conneally, K. E. Schmechel, G. W. Small, A. D. Roses, J. L. Haines, and M. A. Pericak-Vance. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature Genetics*, 7(2):180–184, 1994.
- [40] J. Hardy and D. Allsop. Amyloid deposition as the central event in the aetiology of Alzheimer’s disease. *Trends in Pharmacological Sciences*, 12(10):383–388, 1991.
- [41] K. Beyreuther and C. L. Masters. Amyloid precursor protein (APP) and  $\beta$  A4 amyloid in the etiology of Alzheimer’s disease: Precursor-product relationships in the derangement of neuronal function. *Brain Pathology*, 1(4):241–251, 1991.
- [42] D. J. Selkoe. The molecular pathology of Alzheimer’s disease. *Neuron*, 6(4):487–498, 1991.
- [43] J. A. Hardy and G. A. Higgins. Alzheimer’s disease: The amyloid cascade hypothesis. *Science*, 256(5054):184–185, 1992.
- [44] N. Suzuki, T. T. Cheung, X.-D. Cai, A. Odaka, L. Otvos Jr., C. Eckman, T. E. Golde, and S. G. Younkin. An increased percentage of long amyloid  $\beta$  protein secreted by familial amyloid  $\beta$  protein precursor ( $\beta$ APP717) mutants. *Science*, 264(5163):1336–1340, 1994.
- [45] K. Herrup. The case for rejecting the amyloid cascade hypothesis. *Nature Neuroscience*, 18(6):794–799, 2015.

- [46] J. M. Castellano, J. Kim, F. R. Stewart, H. Jiang, R. B. Demattos, B. W. Patterson, A. M. Fagan, J. C. Morris, G. Kwasi, C. Cruchaga, A. M. Goate, K. R. Bales, M. Steven, R. J. Bateman, and D. M. Holtzman. Human apoE isoforms differentially regulate brain amyloid-beta peptide clearance. *Science Translational Medicine*, 3(89), 2011.
- [47] L. Bertram and R. E. Tanzi. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature Reviews Neuroscience*, 9(10):768–78, 2008.
- [48] E. S. Musiek and D. M. Holtzman. Three dimensions of the amyloid hypothesis: time, space and 'wingmen'. *Nature Neuroscience*, 18(6):800–806, 2015.
- [49] R. J. Bateman, C. Xiong, T. L. S. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, D. S. Marcus, N. J. Cairns, X. Xie, T. M. Blazey, D. M. Holtzman, A. Santacruz, V. Buckles, A. Oliver, K. Moulder, P. S. Aisen, B. Ghetti, W. E. Klunk, E. McDade, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, P. R. Schofield, R. A. Sperling, S. Salloway, and J. C. Morris. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *The New England Journal of Medicine*, 367(9):795–804, 2012.
- [50] T. Gómez-Isla, R. Hollister, H. West, S. Mui, J. H. Growdon, R. C. Petersen, J. E. Parisi, and B. T. Hyman. Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease. *Annals of Neurology*, 41(1):17–24, 1997.
- [51] M. Hutton, C. L. Lendon, P. Rizzu, M. Baker, S. Froelich, H. H. Houlden, S. Pickering-Brown, S. Chakraverty, A. Isaacs, A. Grover, J. Hackett, J. Adamson, S. Lincoln, D. Dickson, P. Davies, R. C. Petersen, M. Stevena, E. De Graaff, E. Wauters, J. Van Baren, M. Hillebrand, M. Joosse, J. M. Kwon, P. Nowotny, L. K. Che, J. Norton, J. C. Morris, L. A. Reed, J. Trojanowski, H. Basun, L. Lannfelt, M. Neystat, S. Fahn, F. Dark, T. Tan-

- nenberg, P. R. Dodd, N. Hayward, J. B. J. Kwok, P. R. Schofield, A. Andreadis, J. Snowden, D. Craufurd, D. Neary, F. Owen, B. A. Costra, J. Hardy, A. Goate, J. Van Swieten, D. Mann, T. Lynch, and P. Heutink. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*, 393(6686):702–704, 1998.
- [52] P. Poorkaj, T. D. Bird, E. Wijsman, E. Nemens, R. M. Garruto, L. Anderson, A. Andreadis, W. C. Wiederholt, M. Raskind, and G. D. Schellenberg. Tau is a candidate gene for chromosome 17 frontotemporal dementia. *Annals of Neurology*, 43(6):815–825, 1998.
- [53] M. G. Spillantini, J. R. Murrell, M. Goedert, M. R. Farlow, A. Klug, and B. Ghetti. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proceedings of the National Academy of Sciences*, 95(13):7737–7741, 1998.
- [54] D. M. Walsh and D. J. Selkoe. A critical appraisal of the pathogenic protein spread hypothesis of neurodegeneration. *Nature Reviews Neuroscience*, 17(4):251–60, 2016.
- [55] C. M. Clark, S. Xie, J. Chittams, D. Ewbank, E. Peskind, D. Galasko, J. C. Morris, D. W. Mckeel, M. Farlow, S. L. Weitlauf, J. Quinn, J. Kaye, D. Knopman, H. Arai, R. S. Doody, C. Decarli, S. Leight, V. M. Lee, and J. Q. Trojanowski. Cerebrospinal fluid tau and beta-amyloid: how well do these biomarkers reflect autopsy-confirmed dementia diagnoses? *Archives of Neurology*, 60(12):1696–702, 2003.
- [56] D. Strozzyk, K. Blennow, L. R. White, and L. J. Launer. CSF A $\beta$  42 levels correlate with amyloid-neuropathology in a population-based autopsy study. *Neurology*, 60(4):652–656, 2003.
- [57] N. S. M. Schoonenboom, W. M. van der Flier, M. A. Blankenstein, F. H. Bouwman, G. J. Van Kamp, F. Barkhof, and P. Scheltens. CSF and MRI

markers independently contribute to the diagnosis of Alzheimer's disease. *Neurobiology of Aging*, 29(5):669–75, 2008.

- [58] M. D. Ikonomovic, W. E. Klunk, E. E. Abrahamson, C. A. Mathis, J. C. Price, N. D. Tsopelas, B. J. Lopresti, S. Ziolko, W. Bi, W. R. Paljug, M. L. Debnath, C. E. Hope, B. A. Isanski, R. L. Hamilton, and S. T. DeKosky. Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. *Brain*, 131(6):1630–45, 2008.
- [59] B. J. Bacskai, M. P. Frosch, S. H. Freeman, S. B. Raymond, J. C. Augustinack, K. A. Johnson, M. C. Irizarry, W. E. Klunk, C. A. Mathis, S. T. Dekosky, S. M. Greenberg, B. T. Hyman, and J. H. Growdon. Molecular imaging with Pittsburgh Compound B confirmed at autopsy: a case report. *Archives of Neurology*, 64(3):431–4, 2007.
- [60] K. Buerger, M. Ewers, T. Pirttilä, R. Zinkowski, I. Alafuzoff, S. J. Teipel, J. DeBernardis, D. Kerkman, C. McCulloch, H. Soininen, and H. Hampel. CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain*, 129(11):3035–41, 2006.
- [61] T. Tapiola, M. Overmyer, M. Lehtovirta, S. Helisalmi, J. Ramberg, I. Alafuzoff, P. Riekkinen, and H. Soininen. The level of cerebrospinal fluid tau correlates with neurofibrillary tangles in Alzheimer's disease. *Neuroreport*, 8(18):3961–3963, 1997.
- [62] K. Blennow and H. Hampel. CSF markers for incipient Alzheimer's disease. *The Lancet Neurology*, 2(10):605–613, 2003.
- [63] J. M. Hoffman, K. A. Welsh-bohmer, M. Hanson, B. Crain, C. Hulette, N. Earl, and R. E. Coleman. FDG PET Imaging in Patients with Pathologically Verified Dementia. *Journal of Nuclear Medicine*, 41(11):1920–1928, 2000.

- [64] C. R. Jack, R. C. Petersen, P. C. O'Brien, and E. G. Tangalos. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology*, 42(1):183–183, 1992.
- [65] C. R. Jack, D. W. Dickson, J. E. Parisi, Y. C. Xu, R. H. Cha, P. C. O'Brien, S. D. Edland, G. E. Smith, B. F. Boeve, E. G. Tangalos, E. Kokmen, and R. C. Petersen. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology*, 58(5):750–757, 2002.
- [66] L. C. Silbert, J. F. Quinn, M. M. Moore, E. Corbridge, M. J. Ball, G. Murdoch, G. Sexton, and J. A. Kaye. Changes in premorbid brain volume predict Alzheimer's disease pathology. *Neurology*, 61:487–492, 2003.
- [67] P. M. Thompson, K. M. Hayashi, G. D. Zubicaray, A. L. Janke, S. E. Rose, J. Semple, D. Herman, M. S. Hong, S. S. Dittmer, D. M. Doddrell, and A. W. Toga. Dynamics of Gray Matter Loss in Alzheimer's Disease. *Journal of Neuroscience*, 23(3):994–1005, 2003.
- [68] M. Maruyama, H. Shimada, T. Suhara, H. Shinotoh, B. Ji, J. Maeda, M.-R. Zhang, J. Trojanowski, V.-Y. Lee, M. Ono, K. Masamoto, H. Takano, N. Sahara, N. Iwata, N. Okamura, S. Furumoto, Y. Kudo, Q. Chang, T. Saido, A. Takashima, J. Lewis, M.-K. Jang, I. Aoki, H. Ito, and M. Higuchi. Imaging of Tau Pathology in a Tauopathy Mouse Model and in Alzheimer Patients Compared to Normal Controls. *Neuron*, 79(6):1094–1108, 2013.
- [69] D. T. Chien, S. Bahri, A. K. Szardenings, J. C. Walsh, F. Mu, M. Y. Su, W. R. Shankle, A. Elizarov, and H. C. Kolb. Early Clinical PET Imaging Results with the Novel PHF-Tau Radioligand [F-18]-T807. *Journal of Alzheimer's Disease*, 34(2):457–468, 2013.
- [70] N. Okamura, S. Furumoto, M. T. Fodero-Tavoletti, R. S. Mulligan, R. Harada, P. Yates, S. Pejoska, Y. Kudo, C. L. Masters, K. Yanai, C. C. Rowe, and V. L. Villemagne. Non-invasive assessment of Alzheimer's

- disease neurofibrillary pathology using 18F-THK5105 PET. *Brain*, 137(6):1762–1771, 2014.
- [71] N. Okamura, R. Harada, S. Furumoto, H. Arai, K. Yanai, and Y. Kudo. Tau PET imaging in Alzheimer’s disease. *Current Neurology and Neuroscience Reports*, 14(11):500, 2014.
- [72] D. M. Holtzman, E. Mandelkow, D. J. Selkoe, D. M. Holtzman, E. Mandelkow, D. J. Selkoe, R. E. Tanzi, F. M. Laferla, K. N. Green, and R. Nixon. Alzheimer Disease in 2020. *Cold Spring Harbor Perspectives in Medicine*, 2(11):a011585, 2012.
- [73] L. S. Schneider, F. Mangialasche, N. Andreasen, H. Feldman, E. Giacobini, R. Jones, V. Mantua, P. Mecocci, L. Pani, B. Winblad, and M. Kivipelto. Clinical trials and late-stage drug development for Alzheimer’s disease: An appraisal from 1984 to 2014. *Journal of Internal Medicine*, 275(3):251–283, 2014.
- [74] C. R. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski. Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade. *The Lancet Neurology*, 9(1):119–28, 2010.
- [75] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, and J. Q. Trojanowski. Tracking pathophysiological processes in Alzheimer’s disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 12(2):207–16, 2013.
- [76] M. Ingelsson, H. Fukumoto, K. L. Newell, J. H. Growdon, E. T. Hedley-Whyte, M. P. Frosch, M. S. Albert, B. T. Hyman, and M. C. Irizarry. Early A $\beta$  accumulation and progressive synaptic loss, gliosis, and tangle formation in AD brain. *Neurology*, 62(6):925–931, 2004.

- [77] R. Craig-Schapiro, A. M. Fagan, and D. M. Holtzman. Biomarkers of Alzheimer's disease. *Neurobiology of Disease*, 35(2):128–40, 2009.
- [78] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.
- [79] P. S. Aisen, R. C. Petersen, M. C. Donohue, A. Gamst, R. Raman, R. G. Thomas, S. Walter, J. Q. Trojanowski, L. M. Shaw, L. A. Beckett, C. R. Jack, W. Jagust, A. W. Toga, A. J. Saykin, J. C. Morris, R. C. Green, and M. W. Weiner. Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimer's & Dementia*, 6(3):239–46, 2010.
- [80] P. Buchhave, L. Minthon, H. Zetterberg, A. K. Wallin, K. Blennow, and O. Hansson. Cerebrospinal fluid levels of  $\beta$ -amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Archives of General Psychiatry*, 69(1):98–106, 2012.
- [81] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoëke, S. L. Macaulay, R. Martins, P. Maruff, D. Ames, C. C. Rowe, and C. L. Masters. Amyloid  $\beta$  deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *The Lancet Neurology*, 12(4):357–67, 2013.
- [82] C. R. Jack, H. J. Wiste, T. G. Lesnick, S. D. Weigand, D. S. Knopman, P. Vemuri, V. S. Pankratz, M. L. Senjem, J. L. Gunter, M. M. Mielke, V. J. Lowe, B. F. Boeve, and R. C. Petersen. Brain  $\beta$ -amyloid load approaches a plateau. *Neurology*, 80(10):890–6, 2013.
- [83] C. R. Jack, P. Vemuri, H. J. Wiste, S. D. Weigand, P. S. Aisen, J. Q. Trojanowski, L. M. Shaw, M. A. Bernstein, R. C. Petersen, M. W. Weiner, and D. S. Knopman. Evidence for ordering of Alzheimer disease biomarkers. *Archives of Neurology*, 68(12):1526–35, 2011.



- [84] C. R. Jack, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. Lowe, K. Kantarci, M. A. Bernstein, M. L. Senjem, J. L. Gunter, B. F. Boeve, J. Q. Trojanowski, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and D. S. Knopman. Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. *Archives of Neurology*, 69(7):856–67, 2012.
- [85] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan. Clinical diagnosis of Alzheimer’s disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease. *Neurology*, 34(7):939–944, 1984.
- [86] B. Dubois, H. H. Feldman, C. Jacova, S. T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O’Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P. J. Visser, and P. Scheltens. Research criteria for the diagnosis of Alzheimer’s disease: revising the NINCDS-ADRDA criteria. *The Lancet Neurology*, 6(8):734–746, 2007.
- [87] B. Dubois, H. H. Feldman, C. Jacova, J. L. Cummings, S. T. Dekosky, P. Barberger-Gateau, A. Delacourte, G. Frisoni, N. C. Fox, D. Galasko, S. Gauthier, H. Hampel, G. A. Jicha, K. Meguro, J. O’Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, M. Sarazin, L. C. de Souza, Y. Stern, P. J. Visser, and P. Scheltens. Revising the definition of Alzheimer’s disease: a new lexicon. *The Lancet Neurology*, 9(11):1118–27, 2010.
- [88] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps. The diagnosis of dementia due to Alzheimer’s disease: recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimer’s & Dementia*, 7(3):263–9, 2011.

- [89] B. Dubois, H. H. Feldman, C. Jacova, H. Hampel, J. L. Molinuevo, K. Blennow, S. T. Dekosky, S. Gauthier, D. Selkoe, R. Bateman, S. Cappa, S. Crutch, S. Engelborghs, G. B. Frisoni, N. C. Fox, D. Galasko, M. O. Habert, G. A. Jicha, A. Nordberg, F. Pasquier, G. Rabinovici, P. Robert, C. Rowe, S. Salloway, M. Sarazin, S. Epelbaum, L. C. de Souza, B. Velas, P. J. Visser, L. Schneider, Y. Stern, P. Scheltens, and J. L. Cummings. Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *The Lancet Neurology*, 13(6):614–629, 2014.
- [90] R. C. Petersen. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3):183–194, 2004.
- [91] P. J. Visser, P. Scheltens, and F. R. J. Verhey. Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease? *Journal of Neurology, Neurosurgery and Psychiatry*, 76(10):1348–1354, 2005.
- [92] G. A. Jicha, J. E. Parisi, D. W. Dickson, K. Johnson, R. Cha, R. J. Ivnik, E. G. Tangalos, B. F. Boeve, D. S. Knopman, H. Braak, and R. C. Petersen. Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Archives of Neurology*, 63(5):674–681, 2006.
- [93] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, P. J. Snyder, M. C. Carrillo, B. Thies, and C. H. Phelps. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):270–9, 2011.
- [94] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack, J. Kaye, T. J. Montine, D. C. Park, E. M. Reiman, C. C. Rowe, E. Siemers, Y. Stern, K. Yaffe, M. C. Carrillo, B. Thies, M. Morrison-Bogorad, M. V. Wagster, and C. H. Phelps. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the Na-

- tional Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):280–92, 2011.
- [95] H. Braak, I. Alafuzoff, T. Arzberger, H. Kretschmar, and K. Del Tredici. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica*, 112(4):389–404, 2006.
- [96] S. S. Mirra, A. Heyman, D. McKeel, S. M. Sumi, B. J. Crain, L. M. Brownlee, F. S. Vogel, J. P. Hughes, G. Van Belle, L. Berg, and Others. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*, 41(4):479, 1991.
- [97] J. Godyn, J. Jonczyk, D. Panek, and B. Malawska. Therapeutic strategies for Alzheimer's disease in clinical trials. *Pharmacological Reports*, 68(1):127–138, 2016.
- [98] H. M. Fonteijn, M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill, S. J. Tabrizi, S. Ourselin, N. C. Fox, and D. C. Alexander. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*, 60(3):1880–9, 2012.
- [99] A. L. Young, N. P. Oxtoby, J. M. Schott, and D. C. Alexander. Data-driven models of neurodegenerative disease. *Advances in Clinical Neuroscience and Rehabilitation*, 14(5):6–9, 2014.
- [100] F. M. LaFerla and K. N. Green. Animal Models of Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, 2(11):a006320, 2012.
- [101] E. D. Roberson, K. Scarce-Levie, J. J. Palop, F. Yan, I. H. Cheng, T. Wu, H. Gerstein, G.-Q. Yu, and L. Mucke. Reducing endogenous tau ameliorates

- amyloid  $\beta$ -induced deficits in an Alzheimer's disease mouse model. *Science*, 316(5825):750–754, 2007.
- [102] N. C. Fox and J. M. Schott. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *The Lancet*, 363(9406):392–394, 2004.
- [103] C. R. Jack, V. J. Lowe, S. D. Weigand, H. J. Wiste, M. L. Senjem, D. S. Knopman, M. M. Shiung, J. L. Gunter, B. F. Boeve, B. J. Kemp, M. Weiner, and R. C. Petersen. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. *Brain*, 132(5):1355–65, 2009.
- [104] R. Y. Lo, A. E. Hubbard, L. M. Shaw, J. Q. Trojanowski, R. C. Petersen, P. S. Aisen, M. W. Weiner, and W. J. Jagust. Longitudinal change of biomarkers in cognitive decline. *Archives of Neurology*, 68(10):1257–66, 2011.
- [105] S. Förster, T. Grimmer, I. Miederer, G. Henriksen, B. H. Yousefi, P. Graner, H.-J. Wester, H. Förstl, A. Kurz, B. C. Dickerson, P. Bartenstein, and A. Drzezga. Regional expansion of hypometabolism in Alzheimer's disease follows amyloid deposition with temporal delay. *Biological Psychiatry*, 71(9):792–7, 2012.
- [106] S. M. Landau, M. A. Mintun, A. D. Joshi, R. A. Koeppe, R. C. Petersen, P. S. Aisen, M. W. Weiner, and W. J. Jagust. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of Neurology*, 72(4):578–86, 2012.
- [107] Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, J. M. Mateos-Perez, and A. C. Evans. Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nature Communications*, 7(5):11934, 2016.
- [108] J. W. Bartlett, C. Frost, N. Mattsson, T. Skillbäck, K. Blennow, H. Zetterberg, and J. M. Schott. Determining cut-points for Alzheimer's disease biomark-

- ers: statistical issues, methods and challenges. *Biomarkers in Medicine*, 6(4):391–400, 2012.
- [109] A. Caroli and G. B. Frisoni. The dynamics of Alzheimer’s disease biomarkers in the Alzheimer’s Disease Neuroimaging Initiative cohort. *Neurobiology of Aging*, 31(8):1263–74, 2010.
- [110] M. R. Sabuncu, R. S. Desikan, J. Sepulcre, B. T. T. Yeo, H. Liu, N. J. Schmansky, M. Reuter, M. W. Weiner, R. L. Buckner, R. A. Sperling, and B. Fischl. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Archives of Neurology*, 68(8):1040–8, 2011.
- [111] A. Schmidt-Richberg, C. Ledig, R. Guerrero, H. Molina-Abril, A. Frangi, and D. Rueckert. Learning Biomarker Models for Progression Estimation of Alzheimer’s Disease. *PloS one*, 11(4):e0153040, 2016.
- [112] R. Guerrero, A. Schmidt-Richberg, C. Ledig, T. Tong, R. Wolz, and D. Rueckert. Instantiated mixed effects modeling of Alzheimer’s disease markers. *NeuroImage*, 2016.
- [113] P. M. Thompson, M. S. Mega, R. P. Woods, C. I. Zoumalan, C. J. Lindshield, R. E. Blanton, J. Moussai, C. J. Holmes, J. L. Cummings, and A. W. Toga. Cortical Change in Alzheimer’s Disease Detected with a Disease-specific Population-based Brain Atlas. *Cerebral Cortex*, 11(1):1–16, 2001.
- [114] R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox. Mapping the evolution of regional atrophy in Alzheimer’s disease : Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, 99(7):1–5, 2002.
- [115] B. C. Dickerson, A. Bakkour, D. H. Salat, E. Feczko, J. Pacheco, D. N. Greve, F. Grodstein, C. I. Wright, D. Blacker, H. D. Rosas, R. A. Sperling, A. Atri, J. H. Growdon, B. T. Hyman, J. C. Morris, B. Fischl, and R. L. Buckner. The cortical signature of Alzheimer’s disease: regionally specific cortical

thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex*, 19(3):497–510, 2009.

- [116] C. R. McDonald, L. K. McEvoy, L. Gharapetian, C. Fennema-Notestine, D. J. Hagler, D. Holland, A. Koyama, J. B. Brewer, and A. M. Dale. Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology*, 73(6):457–465, 2009.
- [117] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. Luthman, J. C. Morris, R. C. Petersen, A. J. Saykin, L. Shaw, L. Shen, A. Schwarz, A. W. Toga, and J. Q. Trojanowski. 2014 Update of the Alzheimer’s Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer’s & Dementia*, 11(6):e1–e120, 2015.
- [118] G. Chetelat, R. La Joie, N. Villain, A. Perrotin, V. De La Sayette, F. Eustache, and R. Vandenberghe. Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer’s disease. *NeuroImage: Clinical*, 2(1):356–365, 2013.
- [119] A. D. Cohen and W. E. Klunk. Early detection of Alzheimer’s disease using PiB and FDG PET. *Neurobiology of Disease*, 72(A):117–122, 2014.
- [120] I. K. Amlie and A. M. Fjell. Diffusion tensor imaging of white matter degeneration in Alzheimer’s disease and mild cognitive impairment. *Neuroscience*, 276:206–215, 2014.
- [121] J. Young, M. Modat, M. J. Cardoso, A. Mendelson, D. Cash, and S. Ourselin. Accurate multimodal probabilistic prediction of conversion to Alzheimer’s disease in patients with mild cognitive impairment. *NeuroImage: Clinical*, 2:735–745, 2013.

- [122] J. Mattila, J. Koikkalainen, A. Virkki, A. Simonsen, M. van Gils, G. Walde-  
mar, H. Soininen, and J. Lötjönen. A disease state fingerprint for evaluation  
of Alzheimer's disease. *Journal of Alzheimer's Disease*, 27(1):163–76, 2011.
- [123] W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, and M. D. Gre-  
icius. Neurodegenerative Diseases Target Large-Scale Human Brain Net-  
works. *Neuron*, 62(1):42–52, 2009.
- [124] J. Zhou, E. D. Gennatas, J. H. Kramer, B. L. Miller, and W. W. Seeley. Pre-  
dicting regional neurodegeneration from the healthy brain functional connec-  
tome. *Neuron*, 73(6):1216–27, 2012.
- [125] A. Raj, A. Kuceyeski, and M. Weiner. A network diffusion model of disease  
progression in dementia. *Neuron*, 73(6):1204–15, 2012.
- [126] A. Raj, E. LoCastro, A. Kuceyeski, D. Tosun, N. Relkin, and M. Weiner.  
Network Diffusion Model of Progression Predicts Longitudinal Patterns of  
Atrophy and Metabolism in Alzheimer's Disease. *Cell Reports*, 10(3):359–  
369, 2015.
- [127] Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, and A. C. Evans. Epidemic  
Spreading Model to Characterize Misfolded Proteins Propagation in Aging  
and Associated Neurodegenerative Disorders. *PLoS Computational Biology*,  
10(11):e1003956, 2014.
- [128] M. C. Donohue, H. Jacqmin-Gadda, M. Le Goff, R. G. Thomas, R. Raman,  
A. C. Gamst, L. A. Beckett, C. R. Jack, M. W. Weiner, J.-F. Dartigues, and  
P. S. Aisen. Estimating long-term multivariate progression from short-term  
data. *Alzheimer's & Dementia*, 10(5):S400–S410, 2014.
- [129] J. W. Ashford and F. A. Schmitt. Modeling the time-course of Alzheimer  
dementia. *Current Psychiatry Reports*, 3(1):20–8, 2001.
- [130] E. Yang, M. Farnum, V. Lobanov, T. Schultz, R. Verbeeck, N. Raghavan,  
M. N. Samtani, G. Novak, V. Narayan, and A. DiBernardo. Quantifying the

- pathophysiological timeline of Alzheimer's disease. *Journal of Alzheimer's Disease*, 26(4):745–53, 2011.
- [131] N. P. Oxtoby, A. L. Young, N. C. Fox, P. Daga, D. M. Cash, S. Ourselin, J. M. Schott, and D. C. Alexander. Learning imaging biomarker trajectories from noisy Alzheimer's disease data using a Bayesian multilevel model. In *Bayesian and Graphical Models for Biomedical Imaging*, volume 8677 of *Lecture Notes in Computer Science*, pages 85–94, 2014.
- [132] B. M. Jerny, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jerny, B. Caffo, and J. L. Prince. A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging Initiative cohort. *NeuroImage*, 63(3):1478–86, 2012.
- [133] M. Bilgel, J. L. Prince, D. F. Wong, S. M. Resnick, and B. M. Jerny. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *NeuroImage*, 134:658–670, 2016.
- [134] A. L. Young, N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain*, 137(9):2564–2577, 2014.
- [135] C. C. Rowe, K. A. Ellis, M. Rimajova, P. Bourgeat, K. E. Pike, G. Jones, J. Fripp, H. Tochon-Danguy, L. Morandau, G. O'Keefe, R. Price, P. Raniga, P. Robins, O. Acosta, N. Lenzo, C. Szoek, O. Salvado, R. Head, R. Martins, C. L. Masters, D. Ames, and V. L. Villemagne. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. *Neurobiology of Aging*, 31(8):1275–83, 2010.
- [136] J. M. Schott, J. W. Bartlett, N. C. Fox, and J. Barnes. Increased brain atrophy rates in cognitively normal older adults with low cerebrospinal fluid A $\beta$ 1-42. *Annals of Neurology*, 68(6):825–34, 2010.



- [137] W. G. Rosen, R. C. Mohs, and K. L. Davis. A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, 141(11):1356–1364, 1984.
- [138] A. Rey. L'examen clinique en psychologie. *Presses Universitaires de France*, 1958.
- [139] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. M.-Y. Lee, and J. Q. Trojanowski. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology*, 65(4):403–13, 2009.
- [140] A. J. Saykin, L. Shen, T. M. Foroud, S. G. Potkin, S. Swaminathan, S. Kim, S. L. Risacher, K. Nho, M. J. Huentelman, D. W. Craig, P. M. Thompson, J. L. Stein, J. H. Moore, L. A. Farrer, R. C. Green, L. Bertram, C. R. Jack, and M. W. Weiner. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's & Dementia*, 6(3):265–73, 2010.
- [141] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–91, 2008.
- [142] K. K. Leung, M. J. Clarkson, J. W. Bartlett, S. Clegg, C. R. Jack, M. W. Weiner, N. C. Fox, and S. Ourselin. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *NeuroImage*, 50(2):516–23, 2010.

- [143] K. K. Leung, J. Barnes, G. R. Ridgway, J. W. Bartlett, M. J. Clarkson, K. Macdonald, N. Schuff, N. C. Fox, and S. Ourselin. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 51(4):1345–59, 2010.
- [144] J. Hardy and D. J. Selkoe. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 297(5580):353–6, 2002.
- [145] U. Andreasson, R. Lautner, J. M. Schott, N. Mattsson, O. Hansson, S.-K. Herukka, S. Helisalmi, M. Ewers, H. Hampel, A. Wallin, L. Minthon, J. Hardy, K. Blennow, and H. Zetterberg. CSF biomarkers for Alzheimer's pathology and the effect size of APOE  $\epsilon 4$ . *Molecular Psychiatry*, 19(2):148–9, 2014.
- [146] C. Jack and D. Holtzman. Biomarker Modeling of Alzheimer's Disease. *Neuron*, 80(6):1347–1358, 2013.
- [147] H. Braak and K. Del Tredici. The pathological process underlying Alzheimer's disease in individuals under thirty. *Acta Neuropathologica*, 121(2):171–81, 2011.
- [148] E. Kok, S. Haikonen, T. Luoto, H. Huhtala, S. Goebeler, H. Haapasalo, and P. J. Karhunen. Apolipoprotein E-dependent accumulation of Alzheimer disease-related lesions begins in middle age. *Annals of Neurology*, 65(6):650–7, 2009.
- [149] K. Blennow, H. Hampel, M. Weiner, and H. Zetterberg. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology*, 6(3):131–44, 2010.
- [150] N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack, and M. W. Weiner. MRI of hippocam-

- pal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain*, 132(4):1067–77, 2009.
- [151] K. K. Leung, J. W. Bartlett, J. Barnes, E. N. Manning, S. Ourselin, and N. C. Fox. Cerebral atrophy in mild cognitive impairment and Alzheimer disease: rates and acceleration. *Neurology*, 80(7):648–54, 2013.
- [152] J. Barnes, O. T. Carmichael, K. K. Leung, C. Schwarz, G. R. Ridgway, J. W. Bartlett, I. B. Malone, J. M. Schott, M. N. Rossor, G. J. Biessels, C. DeCarli, and N. C. Fox. Vascular and Alzheimer's disease markers independently predict brain atrophy rate in Alzheimer's Disease Neuroimaging Initiative controls. *Neurobiology of Aging*, 34(8):1996–2002, 2013.
- [153] J. L. Whitwell, C. R. Jack, J. E. Parisi, D. S. Knopman, B. F. Boeve, R. C. Petersen, T. J. Ferman, D. W. Dickson, and K. A. Josephs. Rates of cerebral atrophy differ in different degenerative pathologies. *Brain*, 130(4):1148–58, 2007.
- [154] J. L. Whitwell, D. W. Dickson, M. E. Murray, S. D. Weigand, N. Tosakulwong, M. L. Senjem, D. S. Knopman, B. F. Boeve, J. E. Parisi, R. C. Petersen, C. R. Jack, and K. a. Josephs. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. *The Lancet Neurology*, 11(10):868–77, 2012.
- [155] J. M. Schott, B. H. Ridha, S. J. Crutch, D. G. Healy, J. B. Uphill, E. K. Warrington, M. N. Rossor, and N. C. Fox. Apolipoprotein E Genotype Modifies the Phenotype of Alzheimer Disease. *Archives of Neurology*, 63(1):155–156, 2006.
- [156] L. A. Beckett. Maximum likelihood estimation in Mallows's model using partially ranked data. In *Probability Models and Statistical Analyses for Ranking Data*, volume 80 of *Lecture Notes in Statistics*, pages 92–107, 1993.

- [157] J. Huang and D. Alexander. Probabilistic Event Cascades for Alzheimer's disease. In *Advances in Neural Information Processing Systems*, pages 3104–3112, 2012.
- [158] N. P. Oxtoby, A. L. Young, D. M. Cash, T. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, and D. C. Alexander. Data-driven models for predicting fine-grained disease progression and symptom onset in dominantly-inherited Alzheimer's disease without reliance upon familial age of onset. *In Preparation*.
- [159] J. C. Morris, P. S. Aisen, R. J. Bateman, T. L. S. Benzinger, N. J. Cairns, A. M. Fagan, B. Ghetti, A. M. Goate, D. M. Holtzman, W. E. Klunk, E. McDade, D. S. Marcus, R. N. Martins, C. L. Masters, R. Mayeux, A. Oliver, K. Quaid, J. M. Ringman, M. N. Rossor, S. Salloway, P. R. Schofield, N. J. Selsor, R. A. Sperling, M. W. Weiner, C. Xiong, K. L. Moulder, and V. D. Buckles. Developing an international network for Alzheimer research: The Dominantly Inherited Alzheimer Network. *Clinical Investigation*, 2(10):975–984, 2012.
- [160] S. M. Landau, M. Lu, A. D. Joshi, M. Pontecorvo, M. A. Mintun, J. Q. Trojanowski, L. M. Shaw, and W. J. Jagust. Comparing positron emission tomography imaging and cerebrospinal fluid measurements of  $\beta$ -amyloid. *Annals of Neurology*, 74(6):826–836, 2013.
- [161] M. L. Schroeter, S. Tiepolt, A. Marschhauser, A. Thöne-Otto, K.-T. Hoffmann, H. Barthel, H. Obrig, and O. Sabri. Dissociation of amyloid biomarkers in PET and CSF in Alzheimer's disease: a case report. *BMC Neurology*, 15:152, 2015.
- [162] A. M. Fagan, C. Xiong, M. S. Jasielec, R. J. Bateman, A. M. Goate, T. L. S. Benzinger, B. Ghetti, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, S. Salloway, P. R. Schofield, R. A. Sperling, D. Marcus, N. J. Cairns, V. D. Buckles, J. H. Ladenson, J. C. Morris, and D. M.

- Holtzman. Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease. *Science Translational Medicine*, 6(226):226ra30, 2014.
- [163] E. M. Reiman, Y. T. Quiroz, A. S. Fleisher, K. Chen, C. Velez-Pardo, M. Jimenez-Del-Rio, A. M. Fagan, A. R. Shah, S. Alvarez, A. Arbelaez, M. Giraldo, N. Acosta-Baena, R. A. Sperling, B. Dickerson, C. E. Stern, V. Tirado, C. Munoz, R. A. Reiman, M. J. Huentelman, G. E. Alexander, J. B. S. Langbaum, K. S. Kosik, P. N. Tariot, and F. Lopera. Brain imaging and fluid biomarker analysis in young adults at genetic risk for autosomal dominant Alzheimer's disease in the presenilin 1 E280A kindred: a case-control study. *The Lancet Neurology*, 11(12):1048–56, 2012.
- [164] A. S. Fleisher, K. Chen, Y. T. Quiroz, L. J. Jakimovich, M. Gutierrez Gomez, C. M. Langois, J. B. S. Langbaum, A. Roontiva, P. Thiyyagura, W. Lee, N. Ayutyanont, L. Lopez, S. Moreno, C. Muñoz, V. Tirado, N. Acosta-Baena, A. M. Fagan, M. Giraldo, G. Garcia, M. J. Huentelman, P. N. Tariot, F. Lopera, and E. M. Reiman. Associations between biomarkers and age in the presenilin 1 E280A autosomal dominant Alzheimer disease kindred: a cross-sectional study. *JAMA Neurology*, 72(3):316–24, 2015.
- [165] T. L. S. Benzinger, T. Blazey, C. R. Jack, R. A. Koeppe, Y. Su, C. Xiong, M. E. Raichle, A. Z. Snyder, B. M. Ances, R. J. Bateman, N. J. Cairns, A. M. Fagan, A. Goate, D. S. Marcus, P. S. Aisen, J. J. Christensen, L. Ercole, R. C. Hornbeck, A. M. Farrar, P. Aldea, M. S. Jasielec, C. J. Owen, X. Xie, R. Mayeux, A. Brickman, E. McDade, W. Klunk, C. A. Mathis, J. Ringman, P. M. Thompson, B. Ghetti, A. J. Saykin, R. A. Sperling, K. A. Johnson, S. Salloway, S. Correia, P. R. Schofield, C. L. Masters, C. Rowe, V. L. Villemagne, R. Martins, S. Ourselin, M. N. Rossor, N. C. Fox, D. M. Cash, M. W. Weiner, D. M. Holtzman, V. D. Buckles, K. Moulder, and J. C. Morris. Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease. *Proceedings of the National Academy of Sciences*, 110(47):E4502–9, 2013.

- [166] D. C. Ryman, P. S. Aisen, T. Bird, A. Danek, N. C. Fox, A. Goate, P. Frommelt, J. B. S. Langbaum, R. Martins, C. L. Masters, R. P. Mayeux, E. McDade, S. Moreno, E. M. Reiman, J. M. Ringman, S. Salloway, J. C. Morris, and R. J. Bateman. Symptom onset in autosomal dominant Alzheimer disease A systematic review and meta-analysis. *Neurology*, 83(3):253–260, 2014.
- [167] A. L. Young, N. P. Oxtoby, S. Ourselin, J. M. Schott, and D. C. Alexander. A simulation system for biomarker evolution in neurodegenerative disease. *Medical Image Analysis*, 26(1):47–56, 2015.
- [168] N. Schuff, D. Tosun, P. S. Insel, G. C. Chiang, D. Truran, P. S. Aisen, C. R. Jack, and M. W. Weiner. Nonlinear time course of brain volume loss in cognitively normal and impaired elders. *Neurobiology of Aging*, 33(5):845–55, 2012.
- [169] A. K. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Sankhya: The Indian Journal of Statistics*, 7(4):401–406, 1943.
- [170] A. L. Young, N. P. Oxtoby, J. Huang, R. V. Marinescu, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander. Multiple Orderings of Events in Disease Progression. In *Information Processing in Medical Imaging*, volume 9123 of *Lecture Notes in Computer Science*, pages 711–722, 2015.
- [171] M. A. Fligner and J. S. Verducci. Distance Based Ranking Models. *JS-TOR: Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):359–369, 1986.
- [172] C. E. Antoniak. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems, 1974.
- [173] M. Meila and H. Chen. Dirichlet Process Mixtures of Generalized Mallows Models. In *Uncertainty in Artificial Intelligence (UAI)*, pages 285–294, 2010.

- [174] M. Meila and L. Bao. Estimation and Clustering with Infinite Rankings. In *Uncertainty in Artificial Intelligence (UAI)*, pages 393–402, 2008.
- [175] M. J. Cardoso, R. Wolz, M. Modat, N. C. Fox, D. Rueckert, and S. Ourselin. Geodesic information flows. In *Medical Image Computing and Computer-Assisted Intervention*, volume 15, pages 262–70, 2012.
- [176] J. D. Rohrer and H. J. Rosen. Neuroimaging in frontotemporal dementia. *International Review of Psychiatry*, 25(2):221–9, 2013.
- [177] J. D. Rohrer, F. Geser, J. Zhou, E. D. Gennatas, M. Sidhu, J. Q. Trojanowski, S. J. Dearmond, B. L. Miller, and W. W. Seeley. TDP-43 subtypes are associated with distinct atrophy patterns in frontotemporal dementia. *Neurology*, 75(24):2204–2211, 2010.
- [178] J. L. Whitwell, C. R. Jack, J. E. Parisi, M. L. Senjem, D. S. Knopman, B. F. Boeve, R. Rademakers, M. Baker, R. C. Petersen, D. W. Dickson, and K. A. Josephs. Does TDP-43 type confer a distinct pattern of atrophy in frontotemporal lobar degeneration? *Neurology*, 75(24):2212–2220, 2010.
- [179] J. L. Whitwell, S. D. Weigand, B. F. Boeve, M. L. Senjem, J. L. Gunter, M. DeJesus-Hernandez, N. J. Rutherford, M. Baker, D. S. Knopman, Z. K. Wszolek, J. E. Parisi, D. W. Dickson, R. C. Petersen, R. Rademakers, C. R. Jack, and K. A. Josephs. Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. *Brain*, 135(3):794–806, 2012.
- [180] J. L. Whitwell, S. A. Przybelski, S. D. Weigand, R. J. Ivnik, P. Vemuri, J. L. Gunter, M. L. Senjem, M. M. Shiung, B. F. Boeve, D. S. Knopman, J. E. Parisi, D. W. Dickson, R. C. Petersen, C. R. Jack, and K. A. Josephs. Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: A cluster analysis study. *Brain*, 132(11):2932–2946, 2009.
- [181] J. Nettiksimmons, D. Harvey, J. Brewer, O. Carmichael, C. DeCarli, C. R. Jack, R. Petersen, L. M. Shaw, J. Q. Trojanowski, M. W. Weiner, and

- L. Beckett. Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. *Neurobiology of Aging*, 31(8):1419–28, 2010.
- [182] J. Nettiksimmons, C. DeCarli, S. Landau, and L. Beckett. Biological heterogeneity in ADNI amnesic mild cognitive impairment. *Alzheimer's & Dementia*, 10(5):511–521, 2014.
- [183] J. Nettiksimmons, L. Beckett, C. Schwarz, O. Carmichael, E. Fletcher, and C. Decarli. Subgroup of ADNI normal controls characterized by atrophy and cognitive decline associated with vascular damage. *Psychology and Aging*, 28(1):191–201, 2013.
- [184] Y. Noh, S. Jeon, J. M. Lee, S. W. Seo, G. H. Kim, H. Cho, B. S. Ye, C. W. Yoon, H. J. Kim, J. Chin, K. H. Park, K. M. Heilman, and D. L. Na. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. *Neurology*, 83(21):1936–44, 2014.
- [185] A. M. Racine, R. L. Kosciak, S. E. Berman, C. R. Nicholas, L. R. Clark, O. C. Okonkwo, H. A. Rowley, S. Asthana, B. B. Bendlin, K. Blennow, H. Zetterberg, C. E. Gleason, C. M. Carlsson, and S. C. Johnson. Biomarker clusters are differentially associated with longitudinal cognitive decline in late midlife. *Brain*, 139(8):2261–74, 2016.
- [186] J. Hwang, C. M. Kim, S. Jeon, J. M. Lee, Y. J. Hong, J. H. Roh, J.-H. Lee, and D. L. Na. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 2:58–67, 2015.
- [187] J. D. Rohrer, J. M. Nicholas, D. M. Cash, J. van Swieten, E. Dopper, L. Jiskoot, R. van Minkelen, S. A. Rombouts, M. J. Cardoso, S. Clegg, M. Espak, S. Mead, D. L. Thomas, E. De Vita, M. Masellis, S. E. Black, M. Freedman, R. Keren, B. J. MacIntosh, E. Rogaeva, D. Tang-Wai, M. C. Tartaglia, R. Laforce, F. Tagliavini, P. Tiraboschi, V. Redaelli, S. Prioni, M. Grisoli,



- B. Borroni, A. Padovani, D. Galimberti, E. Scarpini, A. Arighi, G. Fumagalli, J. B. Rowe, I. Coyle-Gilchrist, C. Graff, M. Fallström, V. Jelic, A. K. Ståhlbom, C. Andersson, H. Thonberg, L. Lilius, G. B. Frisoni, M. Pievani, M. Bocchetta, L. Benussi, R. Ghidoni, E. Finger, S. Sorbi, B. Nacmias, G. Lombardi, C. Polito, J. D. Warren, S. Ourselein, N. C. Fox, and M. N. Rossor. Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: A cross-sectional analysis. *The Lancet Neurology*, 14(3):253–262, 2015.
- [188] A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.
- [189] C. J. Mahoney, J. Beck, J. D. Rohrer, T. Lashley, K. Mok, T. Shakespeare, T. Yeatman, E. K. Warrington, J. M. Schott, N. C. Fox, M. N. Rossor, J. Hardy, J. Collinge, T. Revesz, S. Mead, and J. D. Warren. Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: Clinical, neuroanatomical and neuropathological features. *Brain*, 135(3):736–750, 2012.
- [190] X. Zhang, E. C. Mormino, N. Sun, R. A. Sperling, M. R. Sabuncu, and B. T. T. Yeo. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer’s disease. *Proceedings of the National Academy of Sciences*, pages E6535–E6544, 2016.
- [191] D. P. Hibar, J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivières, N. Jahanshad, R. Toro, K. Wittfeld, ..., N. G. Martin, M. J. Wright, G. Schumann, B. Franke, P. M. Thompson, and S. E. Medland. Common genetic variants influence human subcortical brain structures. *Nature*, 520(7546):224–229, 2015.